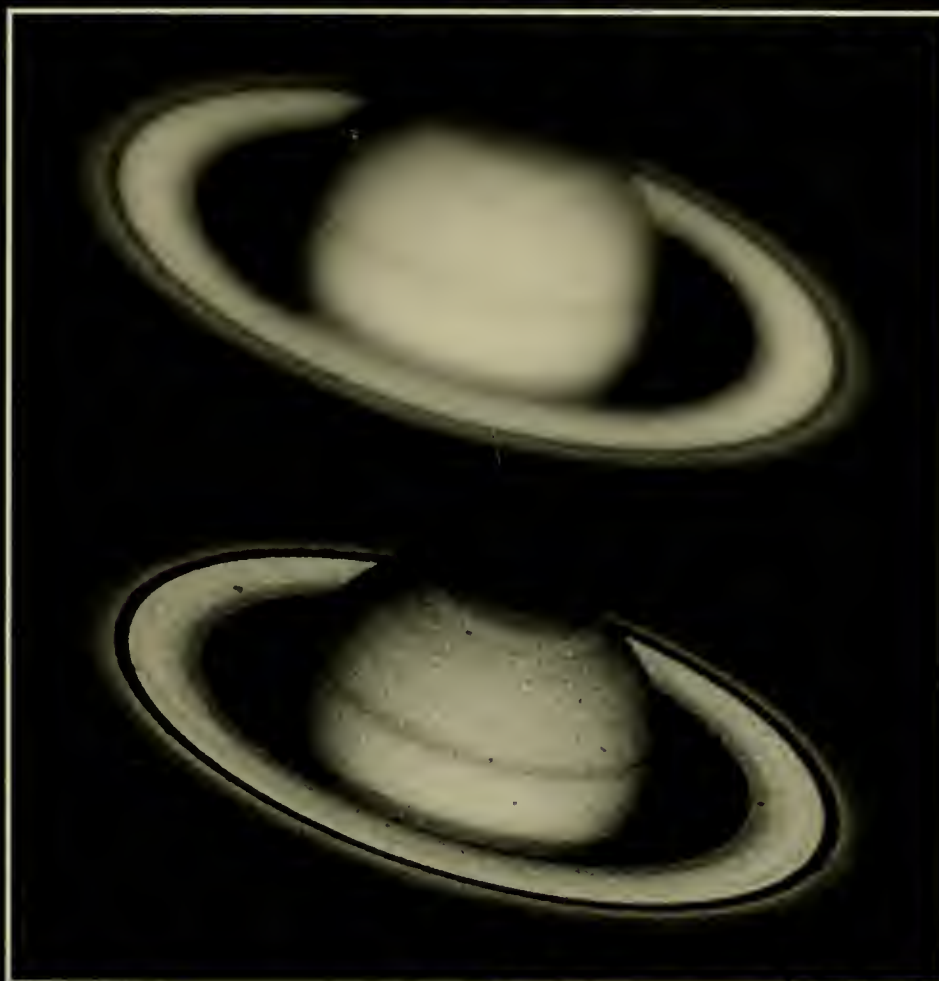# THE RESTORATION OF HST IMAGES AND SPECTRA

Proceedings of a Workshop held at the
Space Telescope Science Institute
Baltimore, Maryland
20-21 August 1990

Edited by R.L. White and R.J. Allen

SPACE
TELESCOPE
SCIENCE
INSTITUTE

NASA
National Aeronautics and
Space Administration

Cover: Original (upper) and restored (lower) image of Saturn. WFC with F 718M filter.
From the paper by Don J. Lindler

SPACE
TELESCOPE
SCIENCE
INSTITUTE

NASA
National Aeronautics and
Space Administration

THE RESTORATION OF HST IMAGES
AND SPECTRA

Proceedings of a Workshop held at the
Space Telescope Science Institute
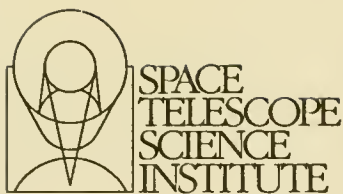Baltimore, Maryland

20-21 August 1990

Editors:

R. L. White and R. J. Allen
Space Telescope Science Institute
Science Computing and Research Support Division

List of Workshop Attendees

John Abbiss
Hans-Martin Adorf
Ron Allen
Scott F. Anderson
David Anthony
Richard Barakat
Eamo Barrett
Frank Bartko
Jim Beletic
Chris Blades
Tj Romke Bontekoe
Bryan Brames
Jack Brandt
Margaret Burbidge
Chris Burrows
Michael Cobb
Greg Cochran
Tim Cornwell
Cindy Cunningham
Chris Dainty
George Djorgovski
Ian Evans
Emilio Falco
Jim Fienup
Don Geman
Riccardo Giacconi
Ron Gilliland
Robert Gonsalves
Joe Goodman
Steve Gull

Bob Hanisch
Richard Harms
Mark Henriksen
Bert Hesselink
Jackie Hewitt
Mike Hollis
Richard Hook
Keith Horne
Joe Horowitz
John Huchra
John Kennealy
J. Kerp
Ed Kibblewhite
Ivan King
Barry Kirkham
Ben Kobler
Rainer Kramm
Anthony Lasenby
Barry Lasker
Tod Lauer
Don Lindler
Jorge Llacer
Leon Lucy
Duccio Macchetto
Mark Maimine
Eliot Malumuth
Stephen P. Maran
Dave Meier
Mark Metzger
Peter Nisensen

Jorge Nunez
Ron Oliversen
Earl O'Neil
Alan Penny
Jorg Pfleiderer
T. Reinheimer
George Reynolds
Mike Roggemann
Ethan Schreier
Jim Schultz
Nigel Sharp
Steve Shore
Susan Simkin
John Skilling
Donald L. Snyder
Brett Spivey
Robert Sprague
Peter Stetson
Owen Storey
Peter Teuben
Charlie Townes
Larry Trafton
Jerzy Usowicz
Wayne Waldron
Archie Warnock
Gerd Weigelt
Nick Weir
Don Wells
Rick White

Table of Contents

Preface

This volume presents the proceedings of the workshop on *The Restoration of HST Images and Spectra*, held at the Space Telescope Science Institute in Baltimore on 1990 August 21–22. The workshop was organized on short notice and was held less than 2 months after the spherical aberration in the Hubble Space Telescope's mirror was discovered. Consequently, relatively little real HST data were available for restoration experiments, and only a few of the workshop participants had access even to that data. Nevertheless, the papers in this volume cover the issues, problems, and techniques quite well and give an indication of directions for future research. Many of the participants have subsequently obtained HST data and have been further studying the problem; we expect that this is only the first in a series of workshops on this topic and that future workshops will have more results of direct relevance to HST.

The papers in these proceedings focus entirely on computational methods for restoring the resolution of HST images and spectra; however, at the workshop we did hear a progress report from Robert A. Brown on proposals to repair the HST hardware. The final report of the HST Strategy Panel has now been issued with conclusions from the work Bob described. Even if HST is equipped with corrective optics for new and existing instruments in the future (as we all hope), computer restoration techniques will still have a useful role to play in the analysis of HST's diffraction-limited data; some examples which easily come to mind are crowded stellar fields and the search for planets around nearby stars.

We have long expected that eventually sophisticated image processing techniques would be applied to HST data; the presence of spherical aberration in HST has pushed us into the restoration game with a vengeance! If there is a bright side to this problem, it is that it may lead astronomers to become more knowledgable about the uses and limits of image restoration methods for a wide range of astronomical data analysis problems.

The organizing committee for this meeting included Ethan Schreier, Chris Blades, and Colin Norman as well as the undersigned. The workshop would never have happened without the able assistance of Barbara Eller, who did a marvelous job of keeping things from falling apart in the face of very short deadlines and a rapidly expanding list of attendees.

Richard L. White
Ronald J. Allen

The HST Spherical Aberration and Its Effects on Images

Richard L. White & Christopher J. Burrows, STScI

The very first image taken a few weeks after the launch of the Hubble Space Telescope (HST) showed evidence for spherical aberration in the HST optics. The results of the next month of testing proved conclusively that the HST primary mirror has about 1/2 wave RMS of spherical aberration ($\lambda = 5000$ Å).

The recently published paper by Burrows *et al.* (1991) gives a detailed description of the HST spherical aberration. In this paper we briefly summarize the problem and discuss its effects on HST imaging science.

The Problem

The HST primary mirror is too flat. The difference Δ between the designed mirror surface and the actual surface varies as $\Delta = 2.3\,\mu\text{m}\ (r/R)^4$, where r is the radial distance from the center of the mirror and $R = 1.2$ m is the radius of the mirror. This error leads to an optical path length error twice as large; the minimum resulting wavefront RMS error is 0.5 waves at $\lambda = 5000$ Å. This error has been determined independently from measurements in orbit and from the flawed ground test equipment which was used to figure the mirror. The measurements currently differ by 10%, but the difference seems to be due to aberrations within the Wide Field/Planetary Camera used to make the measurments, so there is relatively little uncertainty about the nature of problem.

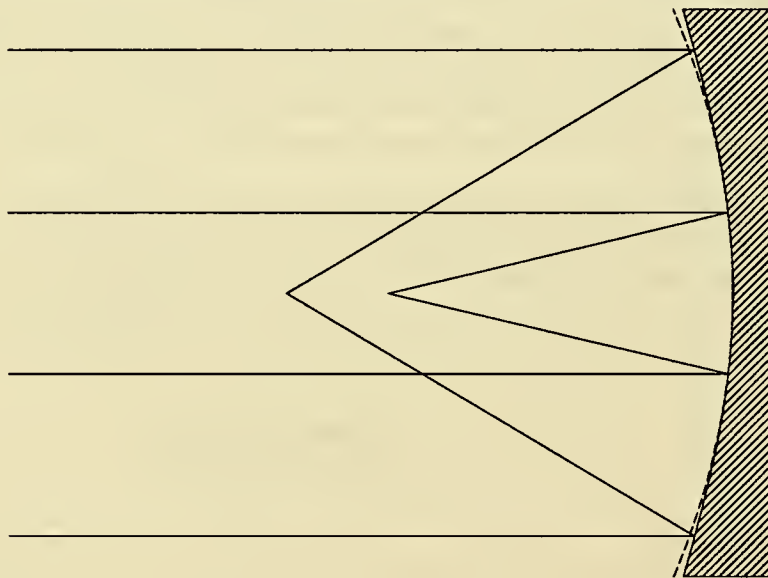Figure 1. Schematic optical diagram showing effect of spherical aberration on paraxial and marginal rays. Desired surface shape is shown with dashed line.

As a consequence of this aberration, light reflected from the center of the HST primary mirror ("paraxial" rays) does not focus at the same point as light reflected from the edge of the mirror ("marginal" rays). The marginal focus is about 4 cm beyond the paraxial

focus. A schematic optical diagram is shown in Figure 1. There is no single best focus for this optical system: as the focus is changed between the paraxial and marginal positions, there is always a annular zone on the mirror for which the rays are focused, but rays inside or outside that zone are unfocused. The resulting images have a small, bright core surrounded by a halo of unfocused light. The adopted focus position was chosen to maximize the fraction of light in the core of the images; this leads to a focus position that is about 1 cm from the paraxial focus.

The HST Point-Spread Function

Figure 2 shows a stellar image obtained close to the best focus setting with the Planetary Camera (PC) through a filter with a central wavelength of 5470 Å. The size of the image core is close to the diffraction limit expected for HST (about 0.1" radius), but contains only about 15% of the energy. The core is surrounded by a halo that encircles the remaining 85% the energy and extends out to a diameter of about 4" from geometrical optics, with light scattered beyond 4" by diffraction effects. The halo is composed primarily of unfocused marginal rays; shadows of obscurations in the pupil such as the four secondary mirror support arms and the three support pads for the primary mirrors are visible in the halo. The fine structure in the halo results from diffraction and varies with wavelength. The point-spread function (PSF) is strongly focus-dependent, and it varies with field position in the Wide Field/Planetary Camera (WF/PC) because of the camera's internal vignetting.

Figure 3 shows encircled-energy and intensity profile plots for the stellar image in Figure 2. As can be seen, 70% of the energy is contained within a radius of about 1.4". (For comparison, the design requirements for the HST optical system called for 70% encircled energy within 0.10" radius.)

Science with HST Images

The images from HST are quite unlike those from ground-based optical telescopes. Ground-based image profiles are determined by turbulence in the atmosphere and are usually roughly Gaussian in shape with diameters of about 1". Most of the energy in such an image is contained within a circle of radius 0.5–1"; consequently, there were some early statements that the HST images are "worse than from the ground."

However, the bright core in the HST PSF is completely lacking in ground-based images. Even though the core contains only 15–20% of the light, its surface brightness is much larger than that of the surrounding halo (Fig. 3). For uncrowded, bright stars the resolution of HST images is comparable to its design goal.

It may be useful to think of HST images as being the sum of *two* images of the same astronomical object: a sharp, well-focused image containing 15% of the light, and a blurry image containing 85% of the light. The sharp image may sometimes be difficult to see against the bright background of the blurry image, but the sharp image is always there if the signal-to-noise ratio (S/N) of the data is sufficient.

The fundamental loss of HST imaging science as a result of the spherical aberration is not a loss of resolution; rather, it is a loss of the ability to detect faint objects, especially in crowded fields. The detection of the faintest sources in astronomical images is limited
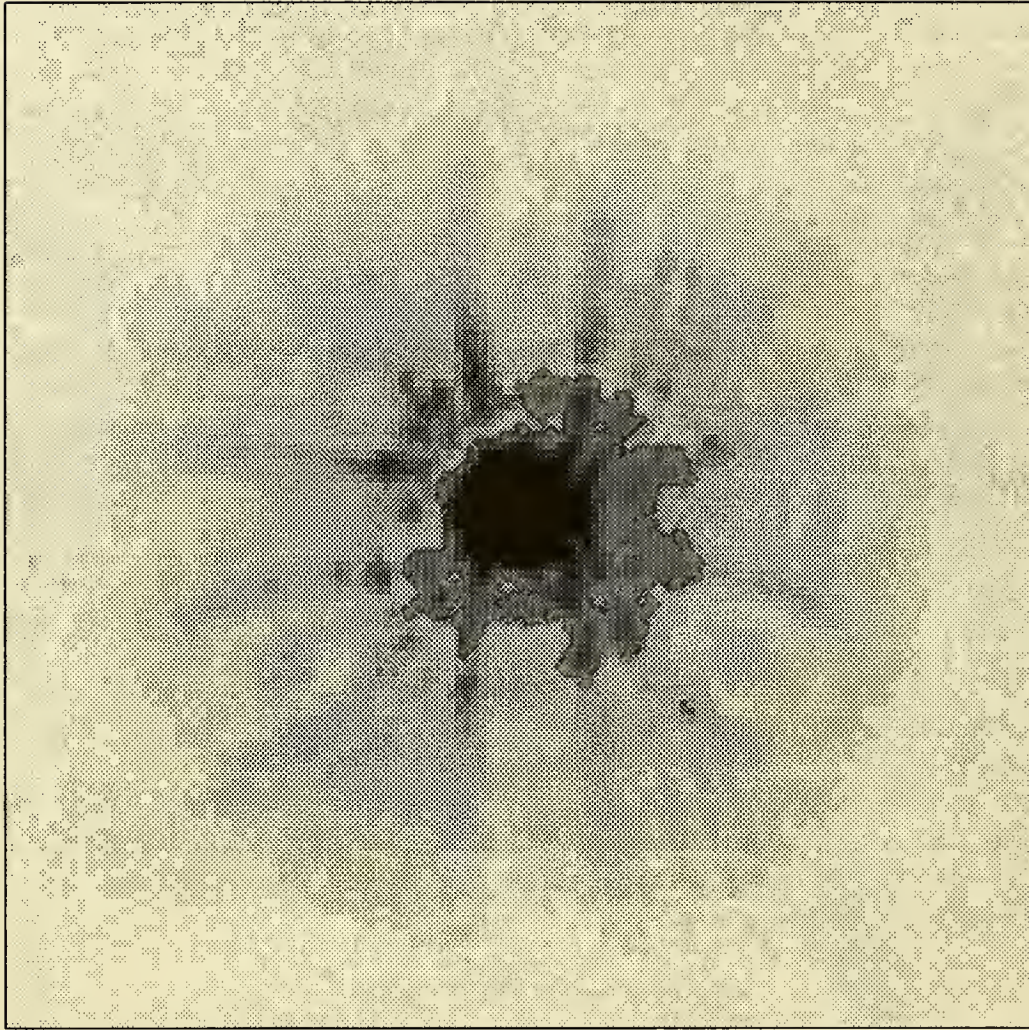
Figure 2. Grey scale representation of a bright star obtained on 15 July 1990 with the Planetary Camera. The field of view of this 200 pixels square sub-image is 8.6 × 8.6 arcsec.
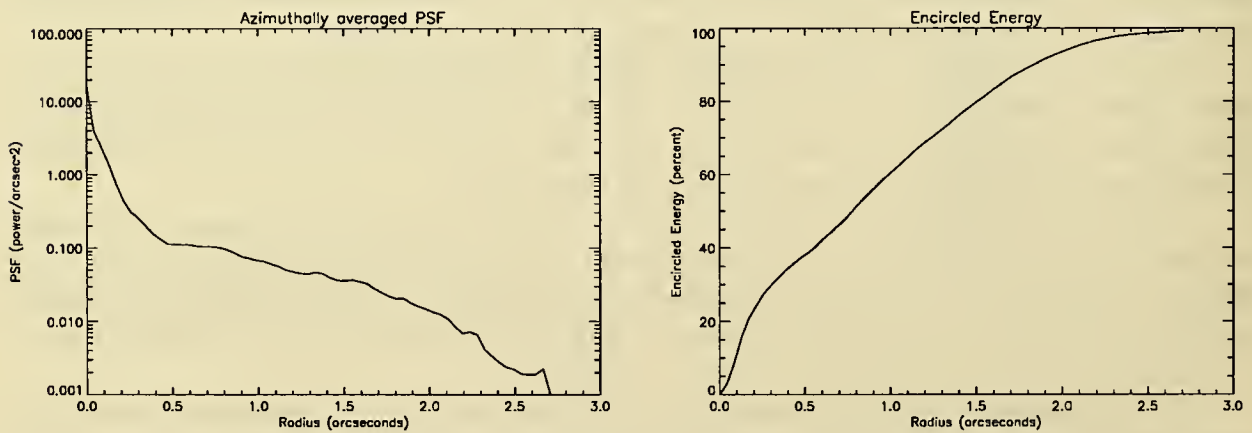

Figure 3. The encircled energy and intensity profile of the star in Figure 2 out to a radius of 4 arcsec.

by the background light from the sky. HST was to have a tremendous advantage over the ground for faint sources because its small images meant that only a very small amount of the sky light need be included when measuring the brightness of faint stars. With the current images, even isolated stars are harder to detect because only 15% of the light is in the PSF core; if one tries to use the extended light in the halo to help detect faint objects, the S/N is actually *reduced* because much more sky light is also included. The situation is much worse in crowded fields, where the overlapping halos of many sources create a much brighter background against which faint objects are even harder to detect.

The end result of all this is that with current HST images the faintest detectable objects are 6 to 10 times brighter than with a perfect HST. No image processing technique can possibly restore these faint objects because the S/N is just not sufficient.

The S/N for brighter stars will obviously also be reduced as a result of the aberration. However, for brighter objects it is possible to increase the exposure time by a factor of a few to recover the lost S/N. For the faintest objects, though, the exposure times were already as long as is practical, so increasing the exposure time is not an option.

Outstanding Problems

We conclude this paper with a brief discussion of what information astronomers want to derive from HST images and what we see as the outstanding problems in HST image restoration. From the early results presented in this volume, it is clear that many image restoration methods are highly successful at deriving images that "look good" from HST data. These restored images may be qualitatively faithful to the true (unknown) image. However, for most astronomical purposes qualitative agreement with reality is not sufficient; we want quantitative agreement as well. For example, it is very important for many scientific programs that we be able to make accurate measurements (to a few percent or better) of the relative brightnesses of stars in crowded fields.

The possible introduction of systematic biases in the brightness of image features during image restoration is consequently of great concern, and the primary open question about restoration of HST images is this: will the restored images be amenable to quantitative analysis or will they be useful as a guide to quantitative analysis of the raw data? For example, it may not be possible to do photometry of crowded stellar fields directly on restored images, but it still may be possible to use the restored image to get a list of star positions, and then to measure the brightness of the stars by fitting PSFs directly to the raw data at those positions.

What is the best way to determine PSFs for HST images? Recall that the PSFs depend on both wavelength (implying a corresponding dependence on stellar color even for a single filter) and on position within the camera field of view for the WF/PC. Empirical calibration of this position and wavelength dependence for all filters and field positions would require an enormous amount of HST observing time. A more attractive possibility is to use optical modeling programs in conjunction with observations to determine theoretical PSFs at wavelengths and positions which have not been observed. A severe complicating factor here is the time-dependence of the focus due to the slow contraction of the HST focal plane structure as the graphite-epoxy outgases. PSF determination is a subject of active research here at STScI.

Even given PSFs which are perfectly known at all positions and wavelengths, the incorporation of these variable PSFs into image restoration methods presents severe computing problems. There are a number of methods which can, in principle, incorporate both color and position dependence; however, it will require orders of magnitude increase in computing power to include the PSF variability for large HST images. Is there a computationally efficient way to incorporate PSF variability?

In conclusion, regardless of the questions still to be answered about the restoration of HST images, it is clear that image restoration will play an important role in the analysis of HST images. If we can solve some of the problems mentioned here, we can expect the results to be widely applicable to other fields; conversely, astronomers should be able to benefit from the research that has already been done outside astronomy on image restoration.

Reference

Burrows, C. J., Holtzman, J. A., Faber, S. M., Bely, P. Y., Hasan, H., Lynds, C. R., and Schroeder, D. 1991, *Ap. J. (Letters)*, in press.

EFFECT OF PSF ON HST SPECTROGRAPHS

Ronald L. Gilliland
Space Telescope Science Institute[1], 3700 San Martin Drive
Baltimore, MD 21218

ABSTRACT Basic capabilities of the two primary HST spectrographs are reviewed. The impact of a poorer than expected PSF enters through loss of throughput for small entrance apertures, loss of spectral resolution for observations through large apertures, and an unavoidable loss of spatial purity for observations in crowded fields. The extent to which deconvolution/restoration techniques can regain lost spectral resolution, and at what cost, for observations through large apertures will be examined.

I. Introduction – the HST spectrographs

The Hubble Space Telescope has two primary spectrographs: the Faint Object Spectrograph (FOS) designed for low spectral resolution observations, and the Goddard High Resolution Spectrograph (GHRS) that will generally be used on brighter objects for high dispersion studies. In addition the Faint Object Camera and Wide Field/Planetary Cameras both offer spectroscopic capabilities. The discussion in this paper is limited to the FOS and GHRS. These two instruments share many similarities, including nearly identical one-dimensional detector systems, allowing a coherent discussion encompassing both.

The FOS offers (see Ford, H.D., and Hartig, G. 1990, *Faint Object Spectrograph Instrument Handbook*, Space Telescope Science Institute, for details) spectral resolving powers, R = $\lambda/\Delta\lambda$, of 1300 and 250 from the far UV at 1100 Å to the far red at 8000 Å. A large variety of spectrograph apertures (slits) are provided to allow varying degrees of spatial isolation.

The GHRS offers (see Duncan, D.K., and Ebbets, D. 1990, *Goddard High Resolution Spectrograph Instrument Handbook*, Space Telescope Science Institute, for details) spectral resolving powers of 100000, 25000, and 2500 in the UV from 1100 to 3200 Å. Only two apertures, the Large Science Aperture (LSA) a 2.0 arcsec square, and the Small Science Aperture (SSA) a 0.25 arcsec square exist for observations. Because the aperture choices for GHRS are quite limited we will concentrate on a discussion of this simple case in the remainder of this paper; the results may be carried over in general terms for the broader range of FOS apertures.

A schematic description of how the spectrographs function is as follows: The telescope optics form an image of the sky (now corrupted by the spherical aberration problem) at the focal plane of

---

HST where the several FOS and GHRS apertures are located. The telescope is positioned to place the target in position relative to the desired aperture. The spectrographs re-image the aperture on a detector system using selectable mirrors (broadband light for target acquisition), or gratings to provide a dispersed light image. Light, as spatially limited on the sky by an aperture, passes to a collimator, is reflected to a grating (or mirror), and thus dispersed onto a light sensitive photocathode. At the photocathode electrons are released, multiplied and accelerated toward a linear array of 500 silicon detector elements (diodes) known as a digicon. The electron beam coming off the photocathode can be magnetically steered, allowing fine control over the positioning of the image on the photocathode relative to the detectors. Each detected photon generates a large number of electrons; the electron pulses are counted individually by incrementing computer memory. As a photon counting device the primary noise term will usually be simple photon counting statistics; at low light levels a dark current of 0.01 counts/s/diode becomes important.

The basic resolution element is defined as one diode width; the spectrum is electronically dithered by 1/2, or 1/4 steps in successive exposures to provide critical sampling. For the GHRS one resolution element is defined exactly by the 0.25 arcsec aperture, the FOS has a corresponding resolution element mapping on the sky of 0.35 arcsec. With spherical aberration a point source now has spatial extent large compared to these resolution domains, therefore observations with apertures larger than 0.25 arcsec, GHRS, or 0.35 arcsec for FOS result in loss of resolving power.

Use of the smaller apertures now result in substantial throughput losses relative to pre-launch predictions; larger apertures allow near-nominal throughput, but serious loss of spectral resolution. Spectral restoration techniques can restore the resolution loss, but only at the expense of increasing noise levels. The primary question to be considered by HST spectroscopists is: Should observations be acquired at high resolution (but low S/N) through small apertures, or obtained through large apertures at better S/N and then deconvolved to restore spectral resolution? (Note that the analogous option does not exist for science observations with the cameras.)

All experiments with simulated data to date suggest that, if the ultimate resolving power is desired, then the observations should be acquired at high resolution (small aperture) initially. This lessens the relative importance of deconvolution techniques for HST spectroscopy. If full restoration of resolving power is not required, it might be advantageous to acquire large aperture observations followed by a partial restoration.

II. PSF Induced throughput and resolution losses

a) Throughput

Figure 1 shows a simulated PSF (at 3400 Å) with the 0.25 and 2.0 arcsec apertures superposed for relative scale. The PSF was generated using software developed by Chris Burrows at ST ScI. The small aperture passes about 15% of the total light, while the large aperture passes 60%, these numbers are smaller than expected by ×4 and ×1.2 respectively. At shorter wavelengths (1200 Å) the relative throughputs are expected to be about 9% and 48% for the SSA and LSA respectively. These numbers will depend on the secondary focus setting ultimately adopted for HST, as well as such details as residual astigmatism of the off-axis aperture position, and primary mirror micro-roughness. On orbit calibrations are not yet available.

8
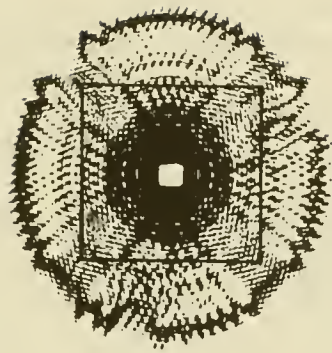
Figure 1. Simulated PSF with spherical aberration. Scale is indicated by superposition of the GHRS apertures, small square at center is 0.25 arcsec, the large box still inside distribution of light represents the 2.0 arcsec aperture.

b) Resolving power

The degradation of spectral resolution may be determined by convolving the distribution of light (along direction of dispersion) with the internal spectrograph line spread function. For both GHRS and FOS the internal line spread function may be approximated as a Gaussian with FWHM = 55μm. (1 arcsec maps to 200 μm, 1 diode width – nominal resolution element – is 50 μm in width.) Figure 2 (upper panel) shows the distribution of light along the dispersion direction, (peak intensity of LSA arbitrarily scaled to 1.0). The lower panel shows convolution of light distribution with internal line spread function. This is what a spatial point and spectral delta-function source is mapped to through the SSA and LSA apertures of GHRS. Relative throughputs are correct as plotted. The FWHM of LSA function is 50% broader than same for SSA. However, the LSA function has broader wings which do not appear with observations through a small aperture. A first moment measure of width shows a resolving power loss of ×2.8 for the LSA relative to SSA. A PSF similar to that shown in lower panel of Figure 2 is what would be used for deconvolution of LSA data.
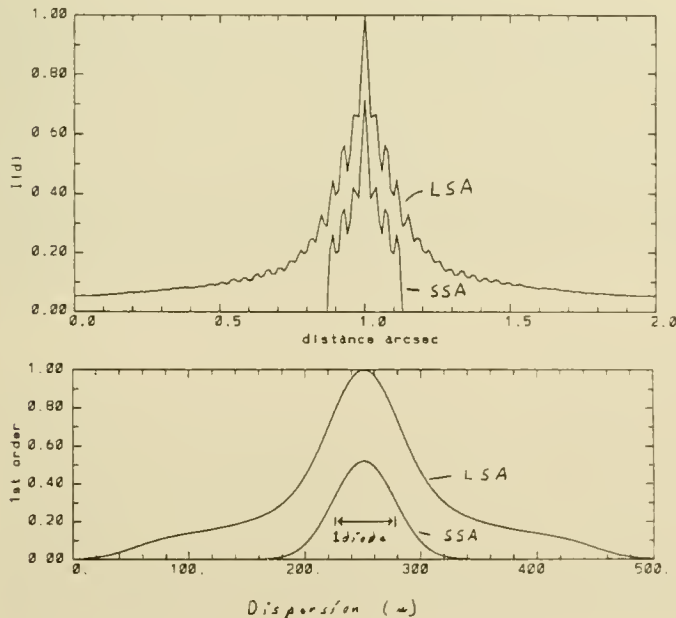
Figure 2. Upper panels show one-dimensional distribution of light along the dispersion direction for GHRS apertures. The ragged structure results from diffraction rings in the PSF, these are smoothed out by the spectrograph's internal line spread function. Lower panels show the spectroscopic PSF; the convolution of light spread (from upper panel) with internal line spread function (a narrow gaussian). The total LSA spread covers 500 μ, or 10 diodes. Relative scales for SSA and LSA are correct.

9

Figure 3 shows a simulation of throughput and resolving power loss. The upper panel is a test spectrum with many sharp absorption lines, without noise. The lower panels show the result of convolution with the PSFs (lower panel of Fig. 2) for SSA and LSA; photon shot noise has then been added. One can see that the S/N of the SSA spectrum is low relative to the LSA simulation, and that the LSA case shows considerable smearing of spectral features.
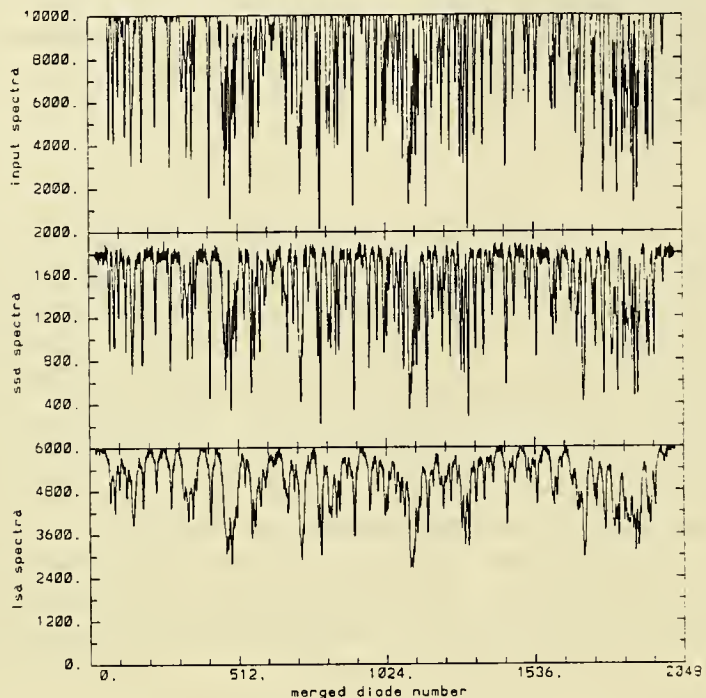
Figure 3. Demonstration of throughput and resolving power losses for the GHRS apertures. Upper panel shows a noiseless input spectrum at nominal continuum level of 10000 counts as would be seen with an aperture admitting 100% of the PSF light and no smearing. Middle panel shows convolution with SSA aperture response function (lowers throughput without loss of resolution), after addition of photon shot noise. Lower panel shows convolution with LSA aperture response function showing the significant loss of resolution.

c) Spatial purity

Some of the highest priority spectroscopic observations with HST required spatial isolation of quite small areas on the sky, e.g., use of FOS and a 0.1 arcsec aperture to observe velocity dispersions in galaxy cores. Since the spatial mixing of light occurs prior to reaching the spectrograph apertures, use of a small aperture is no longer effective at isolating small domains. In order to produce high spatial purity information now, would require acquisition of a spatially distributed raster of individual spectra, and then deducing the contamination of the central point spectrum from surrounding areas. This is a difficult, coupled spatial-spectral deconvolution problem that is beyond the scope of this paper.

III. Spectral restoration of large aperture data

In the spectral restoration of large aperture data to be discussed below, perfect knowledge of the smearing PSF is assumed. This biases results of deconvolution to appear more favorable than may be justified. As with the cameras, adequate measurement of the spatial PSF at the spectrograph apertures will be a difficult task. In principle the spectrographs can form direct images of the large apertures when used in target acquisition mode. For the GHRS at least 3 of 4 target acquisition mirrors introduce spurious structure in the image making the simple determination of the PSF useless. It is possible to measure the PSF directly in dispersed light, if a spatial and spectral point

source emission line object can be used. At the high spectral resolutions of GHRS, and spatial resolution of the HST, few, if any, astronomical PSF calibration sources exist. For the FOS it is more likely that direct measures of light distribution across the aperture can be acquired. Furthermore, with the low resolution of FOS, many astronomical objects have emission line spectra that should allow direct calibration of the total PSF smearing of sharp, unresolved lines. A complication with FOS, which at low resolution covers a large simultaneous spectral domain, is that the PSF will likely vary with position. The FOS gratings will also introduce position dependence to the PSF. Definition of the PSF, as required for successful deconvolutions, will be challenging for both spectrographs.

Let us assume, however, that for test purposes the PSF blurring for observations through the GHRS large science aperture is known. How well can the spectra be deconvolved? Recall that we can directly observe spectra *without* the spectral blurring by use of the SSA. Spacecraft time is the expendable commodity, and (for these purposes) S/N and resolution of spectra the measure of quality. Do the best quality spectra result from LSA observations followed by deconvolution, or simply SSA observations? All simulations to date favor the latter conclusion: Observations that need the resolution offered by the SSA should simply be obtained there. This conclusion is reached despite the over-favorable assumption of a perfectly known PSF. Balancing this would be the caveat that better reconstruction algorithms might change this basic conclusion.

The test simulations have been primarily a reconstruction of the LSA spectrum shown in Figure 3. Four different techniques have been attempted (significant detail is given only for the currently most successful approach):

1) Jansson's (1984). This is a nonlinear, constrained, technique. It's strength is the ability to easily specify both upper and lower solution bounds. Physically a spectrum must be positive, therefore one may certainly require the solution be non-negative. In our simulation (and for many observations) we know that the spectrum is purely absorption and may thus set an upper bound at the (in this case) known continuum level. Relative to an unconstrained technique, such as a Fourier Wiener filter, the placing of bounds on the solution can yield dramatic improvement.

The problem at hand is to solve the integral equation:

$$\phi(x) = \int \psi(x - y)P(y)dy$$

where $\phi(x)$ is the observed spectrum, $P(y)$ is the known point spread function blurring the true spectrum $\psi(x)$ that we desire a robust estimate of. In this notation the Jansson, constrained approach involves the following steps. Iterate the equation:

$$\psi^k(x) = \psi^{k-1}(x) + r^k[\psi^{k-1}(x)][\phi(x) - \int \psi^{k-1}(x)P(y)dy]$$

where the relaxation function r^k is given for example by:

$$r^k[\psi^{k-1}(x)] = r_{max}^k[1 - \mid \psi^{k-1}(x) - (A + B)/2 \mid 2/(B - A)]$$

where r_{max}^k is a constant controlling iterative step size, B is the allowed upper bound (e.g., continuum level), and A is the lower bound (zero enforces positivity). This forces small corrections to be applied near the boundaries.

2) Maximum entropy (Gull and Skilling 1984). This implementation was through the courtesy of Dr. Keith Horne (ST ScI), who had developed a maximum entropy implementation allowing imposition of a bound at the continuum level.

3) Wiener filter (Press, *et al.* 1988). This is simply division in Fourier space of data and PSF transforms multiplied by an optimal filter that goes to zero at high frequencies to suppress noise amplification.

4) Lucy iteration (Lucy 1974). This is a constrained (positivity only in this implementation) nonlinear approach. See the paper by Snyder (1990, this volume) for a detailed discussion of applying this approach to image data.

As a simple measure of reconstruction success we examine the root mean square differences between the original (noiseless) test spectrum and deconvolved LSA and SSA spectra. Table I contains a comparison of the four approaches discussed above.

TABLE I. Deconvolution performance vs. technique

Technique	rms(LSA)	rms(SSA)
Jansson	237.	183.
Maximum Entropy	245.	—
Fourier (Wiener)	313.	215.
Lucy	388.	246.

The Jansson technique, the simplest of all to implement, provides the best result. Although biased toward a favorable result using deconvolution, these simulations have shown that use of the small aperture is preferable to the large, even though the initial S/N is lower by a factor of two. Of course, if the degraded resolution of the LSA is adequate, then it should be used.

IV. Conclusions

The HST spectrographs, unlike the cameras, can acquire data in which the primary impact of the poor PSF is simple loss of light, without any degradation of resolution, through use of narrow slits. Although large slits may yield much more light per unit time than smaller slits, restoration of lost resolution amplifies high spectral frequency noise components beyond that of spectra acquired through small slits. For some specific observations, use of large apertures, followed by deconvolution may be advisable, but planning for such should be supported by detailed simulations.

REFERENCES

Gull, S.F., and Skilling, J. 1984. *MEMSYS User's Manual*, Maximum Entropy Data Consultants Ltd.

Jansson, P.A. 1984. *Deconvolution with Applications in Spectroscopy*, Academic Press, Inc.

Lucy, L. 1974. *Astron. J.*, **79**, 745.

Press, W.H., Flannery, B.P., Teukolsky, S.A., and Vetterling, W.T. 1986. *Numerical Recipes*, Cambridge University Press.

HST IMAGE PROCESSING: DETERMINATION OF BEST FOCUS AND EFFECTS OF PHOTON NOISE

James W. Beletic
Georgia Tech Research Institute, Atlanta, GA 30332

The primary mirror of the Hubble Space Telescope (HST) suffers from spherical aberration that amounts to about 8 waves of wavefront error at the mirror edge. The deleterious effects of this aberration can be partially compensated by defocus. In this paper we present a methodology for optimizing the focus setting for image reconstruction. We find that the best focus is at 5.5 waves of defocus (measured at the edge of the aperture) in the direction opposite to the spherical aberration. This matches the focus setting which optimizes the performance of the high resolution spectrograph. At best focus, the power spectrum transfer function is depressed relative to ideal HST performance by a factor of 30 out to 55% of the diffraction-limit of the telescope; beyond that spatial frequency, the power spectrum transfer function drops to a value of 10^{-4} of ideal. This high frequency attenuation is actually beneficial since it serves as an anti-aliasing filter for the planetary camera which undersamples the image. The ultimate limit to image reconstruction, however, will be photon noise. We derive the signal-to-noise ratio of a photon limited image and present a model for determining the feasibility of imaging complicated extended objects.

I. Introduction

The primary mirror of the Hubble Space Telescope (HST) was incorrectly manufactured such that it suffers from spherical aberration. The mirror is approximately 2 microns thinner at the edge than it should be. The light from the edge of the mirror focuses at a greater distance than does the light from the center of the mirror, i.e. the marginal focus is further than the paraxial focus. This surface error is equivalent to 4 waves at 5500Å (used as standard wavelength); upon reflection, the wavefront error is thus about 8 waves at the edge of the aperture. For the remainder of this paper a value of 8 waves is used for wavefront error; the exact value does not affect the method of analysis, at a later date a more precise aberration value can be used to refine the results.

Spherical aberration is a radially symmetric aberration that obeys a fourth power law. The deleterious effects of the spherical aberration can be somewhat offset by the focus position, however since focus (which is also radially symmetric) affects the wavefront as the second power of the radius it is not possible to exactly compensate for the spherical aberration (see Fig. 1). The point spread functions (PSF's) that result from different focus settings have been presented elsewhere (c.f. Fienberg, 1990). In general, the PSF's have a bright central core that results from the portion of the wavefront that is "in focus" surrounded by a halo which is the unfocused light from the remainder of the wavefront. Changing focus modifies the shape of this

Figure 1 HST Wavefront Error

Spherical aberration produces a radially symmetric wavefront error that is proportional to r^4 where r is the distance from the center of the aperture. Defocus is also radially symmetric but is proportional to r^2 so defocus can be used to compensate but not completely cancel spherical aberration. Shown in the figure is the wavefront resulting from +8 waves of spherical aberration and -5.5 waves of defocus; this focus setting was judged to be the best focus for imaging.

Normalized Aperture Radius
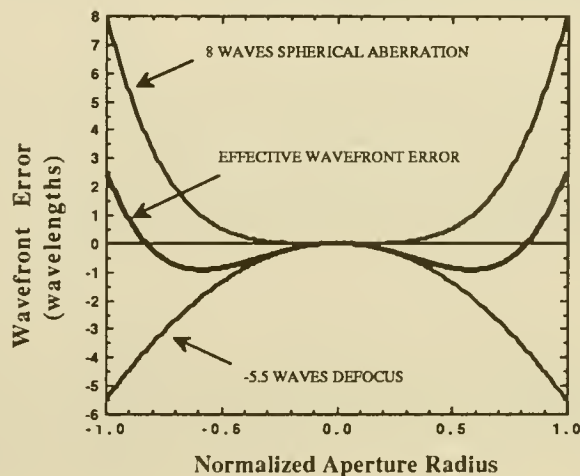
spatial distribution and the best focus setting could depend upon the object which is being viewed or the instrument which is being used.

One criterion for best focus is set by the high resolution spectrograph. An extended object causes the spectrograph to lose wavelength resolution because the spectra of different parts of the object are shifted on the detector relative to each other. This blurs the measured spectrum and fine spectral details may be lost. One method for overcoming the blurring effects of the spherical aberration vis-a-vis the spectrograph is to limit the spatial extent of the object in the entrance slit of the spectrograph. The high resolution spectrograph has two entrance aperture diameters, 0.25 and 1.0 arc sec; using the smaller aperture will minimize the resolution degradation (albeit at the expense of decreased signal level). Thus one criterion for best focus is maximum energy imaged into the 0.25 arc sec aperture. This maximum can be determined from encircled energy curves which are usually measured at a radius 0.10 arc sec. In this paper, we present a straightforward procedure for determining best focus for imaging and compare its results to the encircled energy criterion.

Given the best focus setting for imaging, the reconstructions will still not be ideal. The primary reason for this is photon noise. The effects of photon noise are very difficult to quantify in the image domain, whereas in the spatial frequency domain the photon noise analysis is greatly simplified. We derive the signal-to-noise ratio of photon limited images and present a model that can be used to estimate the feasibility of imaging complicated extended objects.

II. Determination of Best Focus

The best focus for imaging is the focus setting that maximizes the information recorded in the image. The methods of Fourier optics (Goodman, 1968) are appropriate for this investigation. The maximum information will be transmitted by a telescope that maximizes the energy transmitted at all spatial frequencies. For this analysis we make two assumptions:

i)   The objects that the HST will be imaging are *incoherent* emitters, i.e. the light from one part of the object is incoherent with light from any other part of the object. Thus, the image intensity is the sum of the intensities from each part of the object. For astronomical objects, this is a reasonable assumption.

ii) The PSF is the same for all parts of the object, i.e. the PSF is *isoplanatic*. This assumption is in general not valid for the wide field and planetary cameras because the obscuration from the relay secondaries in those cameras varies over the field of view. Non-isoplanatism can not be ignored during image reconstruction unless only a small portion of the field of view is being reconstructed. However for purposes of optimizing the focus we can ignore the effects of non-isoplanatism since the obscuration affects mainly the phase and not the amplitude (i.e. the energy) of the telescope transfer function.

Given these assumptions, the intensity at the image plane, $I(x)$, is given by the convolution of the PSF $P(x)$ and the object $O(x)$,

$$I(x) \; = \; P(x) \star O(x) \tag{1}$$

where x is a two-dimensional vector in the image domain and the symbol $\star$ denotes the convolution operation. Convolution in the image domain is equivalent to multiplication in the spatial frequency domain. We take the Fourier transform of Eq. (1) and follow the convention of normalizing the Fourier quantities (i.e. divide by the DC, or zero frequency, component); normalized quantities are denoted by the tophat symbol ($\hat{\ }$). The spatial frequency vector, $\tilde{\omega}$, is also normalized so that $|\tilde{\omega}| = \omega_{max} = 1$ at the diffraction-limit of the telescope. Thus we can write

$$\hat{I}(\tilde{\omega}) \; = \; \hat{P}(\tilde{\omega})\hat{O}(\tilde{\omega}) \tag{2}$$

The magnitude of the transfer function, $|\hat{P}(\tilde{\omega})|$, is usually referred to as the *modulation transfer function*. The energy is proportional to the magnitude squared of the transfer function, i.e. $|\hat{P}(\tilde{\omega})|^2$. Thus it is useful to express the spatial energy spectrum (or spatial power spectrum) of the image,

$$|\hat{I}(\tilde{\omega})|^2 \; = \; |\hat{P}(\tilde{\omega})|^2|\hat{O}(\tilde{\omega})|^2 \tag{3}$$

The maximum information will be recorded in the image if we maximize the telescope's power spectrum transfer function, $|\hat{P}(\tilde{\omega})|^2$, at all spatial frequencies. The only free parameter in this analysis is the focus setting.

In Fig. 2 the power spectrum transfer function is plotted for the ideal HST, the actual HST, and for computations of +8 waves of spherical aberration with different values of defocus. The defocus values pertain to the number of waves at the mirror edge; the negative values of defocus denote that the focus is opposite the spherical aberration. These curves were calculated with the obscuration appropriate for planetary camera imaging "on-axis", meaning that the relay secondary obscuration is centered in the aperture. These power spectra are calculated as two-dimensional functions but plotted as a function of the magnitude of the spatial frequency, $|\tilde{\omega}|$, in order to obtain a clear picture of the functions.

The general behavior of changing focus setting was found to be as follows. The transfer function is very poor, as would be expected, at paraxial focus. Defocusing in the direction opposite to the spherical aberration enhances the power spectrum of the lowest 50–55% of the spatial frequencies; the enhancement reaches a maximum at a defocus value of -5.5 waves. As defocus is increased beyond -5.5 waves, the mid-range frequencies (5–50%) are attenuated and a peak at 60% of the diffraction-limit appears; the value of this peak reaches maximum at a defocus setting of -7 waves. This peak at $0.6\,\omega_{max}$ is not useful since
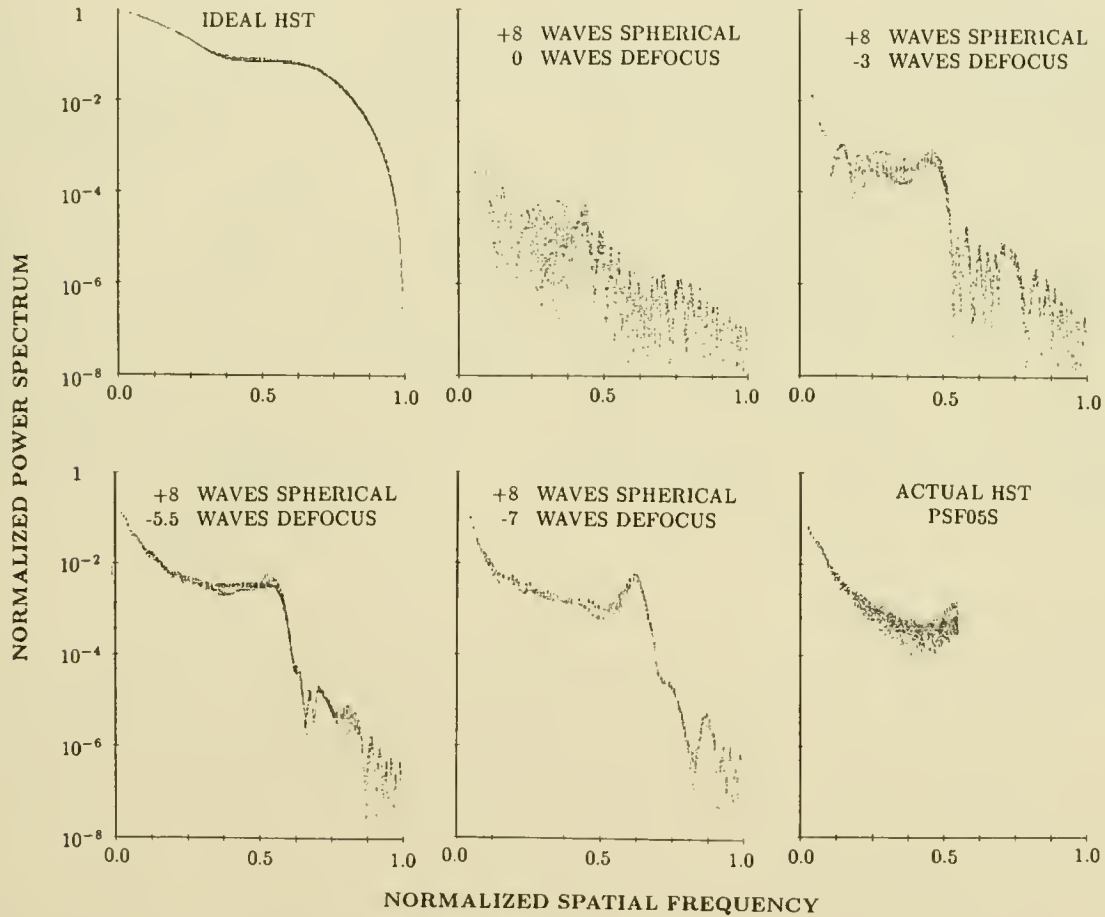
Figure 2 Normalized power spectra, $|\hat{P}(\tilde{\omega})|^2$, as a function of normalized frequency $|\tilde{\omega}|$.

All computations are for planetary camera obscuration (41.4%) and imaging "on-axis", i.e. relay secondary obscuration is centered. The vertical spread in values at a given spatial frequency $|\tilde{\omega}|$ is due to non-symmetry of the obscuration. The actual HST power spectrum is from an early planetary camera image entitled PSF05S, which was taken at focus position labeled "Goddard 0". (Note that all abscissa is linear scale and ordinate is 8 decades of logarithmic scale.)

the wide field (WF) and planetary cameras (PC) undersample the image plane; the pixel sizes of the two cameras are respectively 0.10 and 0.043 arc seconds. (The planetary camera samples to 0.55 ω_{max} at 5500Å.) This undersampling would result in aliasing problems unless the high frequency signal is eliminated prior to detection. In this regard, the transfer function shown for -5.5 waves of defocus is desirable; the power of all spatial frequencies that can be measured by the planetary camera are maximized while higher frequencies are dramatically attenuated, thus avoiding aliasing problems.

The power spectra for -5 and -6 waves defocus are not too different from that shown for -5.5 waves. Thus, the best focus for imaging can be stated as -5.5 ± 0.5 waves of defocus.

Given that a defocus setting of -5.5 waves is close to optimal for imaging, we next wish to determine the best focus for the spectrograph using the 0.10 arc sec encircled energy criterion. Figure 3 presents a plot of the encircled energy at a radius of 0.10 arc sec versus focus setting as computed by our numerical model. Also shown in Fig. 3 is the same curve as measured by the HST. The two curves match very well; both curves peak at about -6.0 waves of defocus. Thus, the best focus for the spectrograph is close enough to the best focus for imaging that one focus setting can be chosen that will satisfy the users of both instruments. This agreement occurs because the telescope transfer function is optimized at all spatial frequencies.

Figure 3 Encircled Energy in 0.10 arc sec radius

The HST curve comes from a presentation dated 7/17/90 by R. Lyon. Any differences between calculated and measured curves are probably due to a few effects that were not taken into account in the numerical calculation. The HST data was taken in the red and measured with a detector that has square area integrating pixels whereas the numerical calculation was done for green light (5500Å) and the PSF was sampled at discrete points. Also, the numerical calculation assumes a perfectly smooth mirror surface with the only wavefront error caused by spherical aberration and defocus. Any microroughness or other aberration (e.g. coma, astigmatism) will cause the HST measurement to have a lower peak value than the numerical computation.
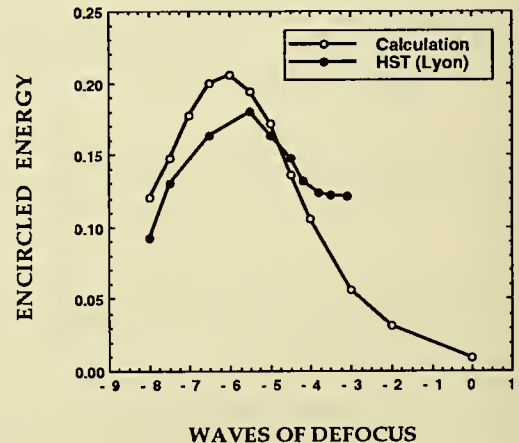
III. Effects of Photon Noise

Over a small portion of the field of view, the intensity distribution that is imaged onto the camera detectors is given by Eq. (1). If we could sense $I(x)$ it would be relatively easy to recover the object $O(x)$ via Fourier inversion; that is, divide both sides of Eq. (2) by $\hat{P}(\tilde{\omega})$ and inverse transform. However, we can never detect the intensity distribution $I(x)$ since energy can only be detected in discrete packets, i.e. photons. In addition, this quantization is the result of a random process that is described by Poisson statistics. The ultimate limit to image reconstruction of the HST images is due to this *stochastic quantization*, which is usually referred to as *photon noise.*

The effects of photon noise on image processing is very difficult to visualize in the image domain. This is because the HST PSF blurs the signal from one object pixel into hundreds of image pixels. Thus the noise at a given image pixel depends not only on the signal from the corresponding object pixel, but also on the object signal from hundreds of neighboring pixels. How can we judge what can be detected and what is beyond the capabilities of image reconstruction? While it is very difficult to get a quantitative handle on photon noise in the image domain, it is relatively straightforward to assess the effects of photon noise in the spatial frequency domain.

The signal-to-noise ratio (SNR) of the spatial power spectrum of a photon limited image (derived in

Appendix A) is given by,

$$\text{SNR}(\tilde{\omega}) \;=\; \frac{\bar{N}^2 |\hat{P}(\tilde{\omega})|^2 |\hat{O}(\tilde{\omega})|^2}{\sqrt{N^2 + 2\bar{N}^3 |\hat{P}(\tilde{\omega})|^2 |\hat{O}(\tilde{\omega})|^2}} \qquad\qquad 0.5 < \tilde{\omega} \le 1 \quad (4)$$

where $\bar{N}$ denotes the number of detected photons. The noise due to limited number of photon events manifests itself as two noise terms; white noise due to limited number of photon events and a signal dependent noise. Eq. (4) applies to the spatial power spectrum, but it has been widely demonstrated that reconstruction of complicated images requires recovery of object phase (c.f. Oppenheim and Lim, 1980). However, the power spectrum SNR provides a very good description of the capability of general image processing algorithms to recover object phase. If no object specific information is employed in the reconstruction algorithms, phase recovery will be limited to those regions where the power spectrum SNR ≥ 1 (Goodman, 1985; Oppenheim and Lim, 1980). Thus, delineation of those regions in the spatial frequency domain where the power spectrum SNR > 1 provides an initial estimate of the capabilities of image reconstruction algorithms.

We make this statement while recognizing the powerful non-linear image reconstruction techniques which employ *a priori* knowledge in the image reconstruction process. If very specific constraints can be applied during the image reconstruction process, then the SNR $= 1$ bound may be too conservative. There are many cases where hard constraints can be applied. For instance: (i) determination of peaked versus smooth intensity distribution in the center of a globular clusters (constrain the answer to one of two distributions), (ii) measurement of the separation of Pluto and Charon (constrain image to consist of two small objects), (iii) determination of the number of point sources in a star cluster (constrain image to consist of a finite number of point sources). These types of constraints are effectively constraints on the knowledge that one desires to obtain from the data.

The power spectrum transfer function, $|\hat{P}(\tilde{\omega})|^2$, was optimized in section II, thus the only parameter in Eq. (4) that can be varied during the imaging process is the number of detected photons, $\bar{N}$. In order to improve SNR, the number of detected photons can be increased by using a wider bandpass filter or lengthening the exposure time. However, there is a limit to the number of photons that can be detected in a given image. This limit is set by the analog-to-digital converter used in the imaging cameras. In the WF/PC, the pixel values are digitized into 12 bits, which corresponds to 4096 possible data values. The conversion factor for the digitizer is set so that one data number corresponds to 7.5 electrons. Thus a maximum value of 4096 corresponds to 30,720 electrons; if more than 30,720 photons are detected in a given pixel, the camera will saturate. Keeping the camera out of saturation limits the total number of detected photons to approximately 30,000 multiplied by the number of pixels that the object is blurred into. For complicated objects, the object spectrum drops so rapidly at high spatial frequencies that even a maximum exposure will not provide good SNR over the whole spatial frequency domain. The signal measured for complicated objects will usually divide into three regions in the spatial frequency domain. The lowest spatial frequencies will have very high SNR and will be easily recovered. The higher spatial frequencies can not be recovered without strong constraints since the SNR is less than 1. Between these two regions is a transition region, over which the SNR falls from a value of 5 (this signal is easily recovered) to a value of 1. Due to the rapid falloff of the object power spectrum of complicated objects, the transition region is often a narrow annulus in the two-dimensional spatial frequency domain.

White noise predictive model

In the general imaging problem, where an unknown field is being examined or a complicated extended object is being imaged, rigid object constraints will not apply. For these "maximum ignorance" imaging scenarios, the power spectrum SNR formula points to a simple model for predicting reconstructed image quality. In these cases, the recovered object signal can be approximated by computing the function $\hat{O}(\tilde{\omega})\hat{P}(\tilde{\omega}) + white\ noise$. The power spectrum of the white noise can be determined from Eq. (4); the mean value is equal to N^{-1} in the normalized power spectrum domain. (When added to $\hat{O}(\tilde{\omega})\hat{P}(\tilde{\omega})$, this white noise should be modeled as a circular complex Gaussian process.) An estimate of the recovered image can be obtained by dividing $\hat{O}(\tilde{\omega})\hat{P}(\tilde{\omega}) + white\ noise$ by $\hat{P}(\tilde{\omega})$ and applying an appropriate low pass filter followed by an inverse Fourier transform. In practice, an image generated in this fashion will probably be of better quality than an actual reconstruction for several reasons. First, without object specific constraints, reconstruction algorithms may have a difficult time achieving phase recovery for signals with power spectrum SNR

$< 2 - 3$. Second, detector noise has not been included in this *white noise model* and signal dependent noise (the second term in the denominator of Eq. 5) has also been ignored. Thus, for an imaging scenario without strong *a priori* object constraints the white noise model will represent a *best case* image reconstruction.

The white noise model can be used to judge the feasibility of certain imaging scenarios. These scenarios include complicated extended objects such as detailed galaxies and the larger planets and Mars. The power spectrum SNR is also useful for any algorithm that needs to establish weights for the different spatial frequency measurements.

When imaging objects for which rigid constraints can be applied, the best reconstruction algorithm will be a very specific one which incorporates the object specific knowledge. The performance will be very sensitive to the application and no general rule of image recovery can be formulated. If the constraints are strong enough it is conceivable that the "limit" set by the power spectrum SNR = 1 will in fact be a *worst case* estimate of image recovery.

Due to space constraints, we do not present white noise model predictions in this paper. We are in the process of simulating various imaging scenarios and comparing images reconstructed by different algorithms with those produced by the white noise predictive model. These efforts will be reported in a forthcoming paper.

IV. Summary

Assuming that the spherical aberration produces +8 waves of wavefront error at the mirror edge, the best focus setting for imaging is -5.5 waves of defocus. This best focus setting is independent of the object which is being imaged and also matches the best focus for the spectrograph. At -5.5 waves of defocus, the telescope's power spectrum transfer function is attenuated by about a factor of 30 out to 55% of the diffraction-limit and by a factor of 10^{-4} at higher spatial frequencies. The attenuation at high spatial frequencies is beneficial since it serves the purpose of an ideal anti-aliasing filter for the planetary camera. The wide field camera will still suffer from aliasing; this effect should be addressed by the image reconstruction algorithms.

The effect of photon noise on images, which is difficult to visualize and quantify in the image domain, is most easily estimated by computing the spatial power spectrum SNR. The power spectrum SNR can be used in any reconstruction algorithm which requires weighting of spatial frequency domain estimates; this function also provides a very good estimate of those frequencies that can be recovered by general image processing algorithms. The power spectrum SNR leads to a simple model for predicting reconstructed image quality, the *white noise model*. However the white noise model will not be appropriate if strong object specific constraints can be applied during the reconstruction process.

Appendix A. Signal-to-Noise Ratio of Photon Limited Images

A photon-limited image is an image which is the composition of a finite number of detected photons. In this appendix we derive the signal-to-noise ratio (SNR) of the spatial power spectrum of a photon-limited image. We do not include the effect of detector noise, which in the case of the HST WF/PC camera is readout noise of about 13 e^- per pixel. Previous analyses of photon-limited imaging (Goodman and Belsher, 1976; Dainty and Greenaway, 1979; Roddier, 1986) have investigated the SNR of the spatial power spectrum when the image intensity is a random variable. That case, where the photon statistics are compounded by a fluctuating intensity level, is called a *compound Poisson process*. Our task is simpler; we only consider statistics of photon detection since the underlying classical image intensity is constant. Our derivation

employs the same notation as used in the aforementioned papers; we closely follow the approach in Roddier (1986).

The image $d(x)$ comprises a series of delta functions which represent detected photons

$$d(x) \; = \; \sum_{k=1}^{N} \delta(x - x_k) \tag{A.1}$$

where x_k denotes the location of the kth photon event and N is the total number of detected photons; N is a random variable. The detected photon events result from a Poisson process that has a rate $\lambda(x)$ proportional to the classical intensity. We wish to determine the SNR of the bias compensated spatial power spectrum. Denoting the Fourier transforms of $d(x)$ and $\lambda(x)$ as $D(\nu)$ and $\Lambda(\nu)$, the bias compensated spatial power spectrum, $Q(\nu)$, and its expectation value, $\mathrm{E}\,[Q(\nu)]$, are (Dainty and Greenaway, 1979):

$$Q(\nu) \; = \; |D(\nu)|^2 - N \tag{A.2}$$

$$\mathrm{E}\,[Q(\nu)] \; = \; |\Lambda(\nu)|^2 \tag{A.3}$$

The expectation operator $\mathrm{E}[\;]$ is used to denote averages over photon statistics. SNR is defined as the mean value divided by the standard deviation, σ, of the value.

$$\mathrm{SNR}(\nu) \; = \; \frac{\mathrm{E}\,[Q(\nu)]}{\sigma\,(Q(\nu))} \tag{A.4}$$

From Eq. (A.3) we can write the variance, σ^2, as,

$$\begin{aligned} \sigma^2\,(Q(\nu)) \; &= \; \mathrm{E}\,[Q^2(\nu)] - (\mathrm{E}\,[Q(\nu)])^2 \\ &= \; \mathrm{E}\,[Q^2(\nu)] - |\Lambda(\nu)|^4 \end{aligned} \tag{A.5}$$

From Eq. (A.2) we get,

$$\mathrm{E}\,[Q^2(\nu)] \; = \; \mathrm{E}\,[|D(\nu)|^4] - 2\mathrm{E}\,[N|D(\nu)|^2] + \mathrm{E}\,[N^2] \tag{A.6}$$

The first term on the right hand side of Eq. (A.6) is calculated in Appendix A of Roddier (1986).

$$\begin{aligned} \mathrm{E}\,[|D(\nu)|^4] \; = \; & \bar{N} + \bar{N}^2 + (\bar{N} + 2)(2|\Lambda(\nu)|^2) + |\Lambda(0)|^2 + |\Lambda(\nu)|^4 + |\Lambda(2\nu)|^2 \\ & + 2\mathrm{Re}\,[\Lambda(2\nu)\Lambda^*(\nu)\Lambda^*(\nu)] + 2\mathrm{Re}\,[\Lambda(0)\Lambda^*(\nu)\Lambda(\nu)] \end{aligned} \tag{A.7}$$

where the symbol * denotes complex conjugation and $\mathrm{Re}[\;]$ denotes the real part of a complex quantity. Since $\bar{N} = \Lambda(0)$ is the average number of photons in a frame, Eq. (A.7) can be simplified to

$$\begin{aligned} \mathrm{E}\,[|D(\nu)|^4] \; = \; & \bar{N} + 2\bar{N}^2 + 4(\bar{N} + 1)|\Lambda(\nu)|^2 + |\Lambda(\nu)|^4 \\ & + |\Lambda(2\nu)|^2 + 2\mathrm{Re}\,[\Lambda(2\nu)\Lambda^*(\nu)\Lambda^*(\nu)] \end{aligned} \tag{A.8}$$

The second and third terms on the right hand side of Eq. (A.6) are calculated in Dainty and Greenaway (1979),

$$\mathrm{E}\,[N|D(\nu)|^2] \; = \; \bar{N} + \bar{N}^2 + (\bar{N} + 2)|\Lambda(\nu)|^2 \tag{A.9}$$

$$\mathrm{E}\,[\bar{N}^2] \; = \; \bar{N} + \bar{N}^2 \tag{A.10}$$

Putting Eqs. (A.8)–(A.10) into Eq. (A.6) and (A.5) we get the variance of $Q(\nu)$,

$$\begin{aligned} \sigma^2\,(Q(\nu)) \; = \; & \bar{N}^2 + 2\bar{N}|\Lambda(\nu)|^2 \\ & + |\Lambda(2\nu)|^2 + 2\mathrm{Re}\,[\Lambda(2\nu)\Lambda^*(\nu)\Lambda^*(\nu)] \end{aligned} \tag{A.11}$$

Equation (A.11) exhibits the "half-frequency" phenomenom which is a fundamental property of photon-limited images (Goodman, 1985). The signal at spatial frequency 2ν induces fluctuations of the signal at spatial frequency ν. If we look at frequencies ν that are greater than half of the highest spatial frequency the signal is zero at 2ν and the SNR is,

$$\text{SNR}(\nu) = \frac{|\Lambda(\nu)|^2}{\sqrt{\bar{N}^2 + 2\bar{N}|\Lambda(\nu)|^2}} \qquad 0.5 < \nu \leq 1 \quad \text{(A.12)}$$

Integration of the spatial power spectrum from M frames of photon-limited data will improve SNR by the factor $\sqrt{M}$; if a series of images can be registered and co-added in the image domain the SNR will improve by the factor M. Using the normalized quantities (denoted by the tophat, $\hat{\ }$) for the object spatial power spectrum, $|\hat{O}(\tilde{\omega})|^2$, the square of the modulation transfer function, $|\hat{P}(\tilde{\omega})|^2$, and the normalized spatial frequency variable $\tilde{\omega}$ we have $|\Lambda(\tilde{\omega})|^2 = \bar{N}^2|\hat{P}(\tilde{\omega})|^2|\hat{O}(\tilde{\omega})|^2$ and can write,

$$\text{SNR}(\tilde{\omega}) = \frac{\bar{N}^2|\hat{P}(\tilde{\omega})|^2|\hat{O}(\tilde{\omega})|^2}{\sqrt{\bar{N}^2 + 2\bar{N}^3|\hat{P}(\tilde{\omega})|^2|\hat{O}(\tilde{\omega})|^2}} \qquad 0.5 < \tilde{\omega} \leq 1 \quad \text{(A.13)}$$

In the high light level limit where $\bar{N}|\hat{P}(\tilde{\omega})|^2|\hat{O}(\tilde{\omega})|^2 \gg 1$,

$$\text{SNR}(\tilde{\omega}) = \sqrt{\frac{\bar{N}}{2}}|\hat{P}(\tilde{\omega})||\hat{O}(\tilde{\omega})| \qquad 0.5 < \tilde{\omega} \leq 1 \quad \text{(A.14)}$$

SNR is proportional to the square root of the number of detected photons. For very low light levels where $\bar{N}|\hat{P}(\tilde{\omega})|^2|\hat{O}(\tilde{\omega})|^2 \ll 1$,

$$\text{SNR}(\tilde{\omega}) = \bar{N}|\hat{P}(\tilde{\omega})|^2|\hat{O}(\tilde{\omega})|^2 \qquad 0.5 < \tilde{\omega} \leq 1 \quad \text{(A.15)}$$

SNR is proportional to the number of detected photons.

Equations (A.12)–(A.15) strictly apply for spatial frequencies $0.5 < \tilde{\omega} \leq 1$. However, since the spatial power spectra of complicated objects decrease with higher spatial frequency, Eqs. (A.12)–(A.15) can often be used for calculating SNR at all spatial frequencies.

References

Dainty, J.C., and A.H. Greenaway (1979). Estimation of spatial power spectra in speckle interferometry. *J. Opt. Soc. Am.* **69**, 786-790.

Fienberg, R.T. and R.W. Sinnott (1990). Space Telescope: Picking up the pieces. *Sky and Telescope*, Vol. 80, No. 4, 352-358.

Goodman, J.W. (1968). *Introduction to Fourier Optics*, McGraw-Hill, New York.

Goodman, J.W. (1985). *Statistical Optics*, Wiley, New York.

Goodman, J.W., and J.F. Belsher (1976). Fundamental limitations in linear invariant restoration of atmospherically degraded images. *SPIE Proc.* **75**, 141-154.

Oppenheim, A. and J. Lim (1981). The importance of phase in signals. *Proc. IEEE* **69**, 529-541.

Roddier, F. (1986). Pupil plane versus image plane in Michelson stellar interferometry. *J. Opt. Soc. Am. A* **3**, No. 12, 2160-2166.

Phase Retrieval Using HST Images

A. N. Lasenby

Mullard Radio Astronomy Observatory,
Cavendish Laboratory,
Madingley Road,
Cambridge CB3 OHE, U.K.

1   Introduction

This workshop is aimed primarily at the questions of what are the best ways of reconstructing HST images, and what is the real degradation in performance of the HST that the spherical aberration has caused. However, in order to carry out the image reconstruction, we first need to know the true form of the aberration, and hence the point spread function (psf) in as much detail as possible. With the telescope already in orbit, the most practical way of doing this is via the images of stars (i.e. assumed point sources) themselves. Thus, an exercise in 'phase retrieval' is necessary, in which we attempt to use the intensity-only data available in the far field of the telescope to reconstruct both phase (i.e. the aberrations) and amplitude (the pupil blockage) over the surface of the mirror. For this purpose, the ideal images required are those where the telescope has been deliberately defocussed by large amounts (equivalent to several wavelengths at the mirror edge) so that imperfections in the surface, and the form of the pupil blockage function, can have a large effect upon the measured data. This talk discusses a new algorithm for achieving this phase reconstruction from out-of-focus images, and shows applications to both an exisiting millimetre telescope in Hawaii, and the HST itself.

2   Why do we need phase retrieval?

Although it was not known at the workshop itself, at the time of writing-up this talk it is thought that the probable cause of the spherical aberration seen in HST images is a mis-grinding of the primary mirror, caused by a specific fault in the null corrector used in the grinding process. The tracing of this fault means that the probable surface to which the mirror was ground can be inferred, and (via a forward transform) the form of the telescope beam observed on point sources can be predicted. This has apparently led to agreement with the measured psf's at about the 10% level (Burrows, p.c.). In this case it might be wondered why the inverse exercise, of deriving the mirror shape from the measured point spread functions (which because of the lack of phase information cannot be carried out simply via an inverse transform) should be necessary. Also, for deconvolution purposes it might be thought that empirical psf's, deduced from neighbouring stars in the field, might be quite adequate for deconvolving extended objects, so that analytically constructed psf's would not be necessary at all. However, in order to make *accurate* deconvolutions, and in some cases to make the deconvolution process work at all, it is necessary to have a point spread function that is as *noise-free* as possible and resembles the actual beam of the telescope, and its variation over the field, as closely as possible. Also, since we cannot be 100% sure that the fault found in the null corrector was the only fault affecting the making of the optical components of the telscope, it makes sense to try to use all the information available to ascertain the true distortion of the incoming wavefronts at the aperture plane, and out-of-focus stellar images provide the most direct route to this. Other motivations for this work include the following:

1. In carrying out the phase reconstruction, accurate physical information about the geometry and diffraction properties of the total instrument have to be incorporated. Once a map of the mirror surface has been generated, using a variety of test images, the same information used in a forward

transform means that psf's can be easily generated for any frequency, filter bandwidth, camera, secondary mirror offset and field position. In particular, the highly variable psf of the Wide Field Planetary Camera (WFPC) can be modelled, and one of the aims of the present work is to be able to infer some of the details of the variable internal obscuration inside the WFPC responsible for this, using the out-of-focus stellar images themselves, since full information is not currently available from pre-flight documents, and in any case depends upon the form of the gross aberrations.

2. The images could be used to find the actual tilt and decentre of the secondary mirror relative to the optimum. This is possible because the phase terms induced by secondary tilt and decentre have a specific recognizable form which can be fitted for as part of the overall process of finding the aperture plane phase front. Thus instead of having to search for this optimum position, the moves necessary to achieve it could be calculated directly from the data.

3. Similar comments apply to the primary mirror actuators. A detailed map of the primary mirror surface would enable the absolute positions of the actuators to be found accurately, and hence the optimum movements to be carried out.

4. Obviously, a very important reason for finding the form of the optical errors accurately is so that the corrective optics for the instrument refits due in 1993 can be prescribed properly.

5. Finally, this work is important simply on the basis that we might see something we did not expect, for example in the way of a blockage or vignetting that was not predicted, or some features of the primary and secondary mirrors that are not expected from the manufacturing procedures as they are currently understood. An example of this is given for the millimetre telescope in Hawaii, on which the current technique has already been tested.

3 The Method

As stated above, the basic problem to be faced for the HST is that the measurements in the far-field of the telescope are of amplitude only (actually amplitude squared), rather than amplitude and phase. If we did have phase available, then we could use the standard scalar aperture integral result

$$E(x, y) = \int \int e^{-i2\pi(ux+vy)} e^{i2\pi(u^2+v^2)/(2R/\lambda)} F(u, v) \, du \, dv,$$

to infer the surface phase $\phi(u, v)$ via a straightforward Fourier transform. Here $E(x, y)$ is the complex beam pattern in amplitude and phase, $F(u, v) = A(u, v) \exp(i\phi(u, v))$ is the surface illumination function, λ is the wavelength of observation and R the distance to the source. u and v are measured in units of wavelength and x and y are (small) angles, in radians. Note that throughout this text, the terms aperture plane and beam plane are taken to refer to the (u, v) and (x, y) planes respectively.

Typically the phase of $E(x, y)$ can be measured with an interferometer system and this technique has in fact been used successfully to set the surfaces of several radio and millimetre telescopes (see Lasenby, 1985, and Padin, Davis & Lasenby, 1987, for two examples). The interferometer systems required can be expensive and complicated to set up, however, and alternative methods, using $|E(x, y)|$ only, have been under development over recent years (see Morris, 1985, for an early review of some of these). The key component of these methods is that more than one amplitude map is taken, with at least one of these being defocussed, and then these are used either in sequence or simultaneously to reconstruct a model of the telescope surface. The defocussing means that irregularities in the telescope surface and the pattern of the blocking function have a greater effect upon the data, and taking data at *different* defocussed positions means that some of the constraints lost due to the missing phase information are resupplied by independent amplitude points.

The method used most commonly up to now, is the Misell method, borrowed from electron microscopy, and used by Jaap Baars and Dave Morris of IRAM to measure the surface of the IRAM 30 m dish at
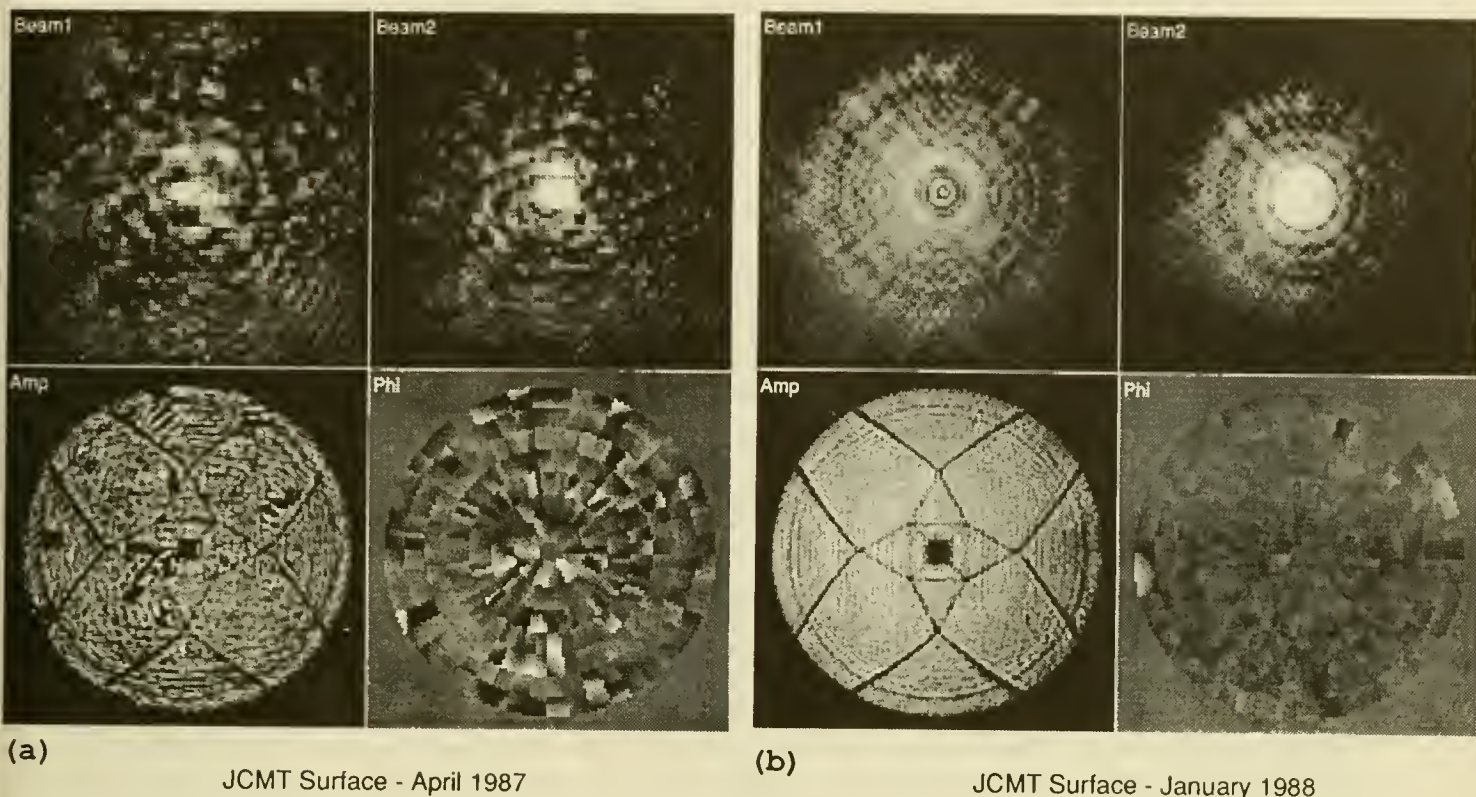
(a) JCMT Surface - April 1987

(b) JCMT Surface - January 1988

Figure 1: *(a) The defocussed beams and derived surface amplitude and phase for the James Clerk Maxwell Telescope in its inital state of adjustment. The beam patterns are built up by raster scanning the telescope across a terrestrial transmitter at 94 GHz. Note that the phase shown has been fitted to the arrangement of surface panels. The incoherent appearance of the beams is a reflection of the disordered state of the panel adjustments. (b) The same but after several rounds of measurement and adjustment using phase retrieval (with the later rounds using large scale fitting).*

Granada (Baars & Morris, 1988) and more recently the 15 m ESO/IRAM/Onsala millimetre dish in Chile (Whybourne & Morris, p.c.). In the Misell method, beams are measured at *two* focus positions, one defocussed and one in focus. An initial guess is made of the amplitude and phase over the surface, and this is then forward transformed to produce a synthetic beam, in amplitude and phase at the first focus position. The measured amplitude is then substituted for the predicted one, but keeping the predicted phase, and an inverse transform carried out, followed by a forward transform to the other beam plane, where the real amplitude measured here is substituted and an inverse transform carried out back to a new surface. This process then continues iteratively until an equilibrium is reached.

In 1987, Richard Hills, Devinder Sivia and myself at MRAO Cambridge (later joined by Dave Waymont), at first used a variant of this Misell method to try to set the surface of the newly completed UK-Dutch-Canada James Clerk Maxwell Telescope sited at the summit of Mauna Kea, Hawaii. The JCMT is a 15 m plain Cassegrain system, with 276 independently movable surface panels, intended for mm/sub-mm spectral line and continuum work. The panels have a surface accuracy of $\sim 10\,\mu$m, and it was hoped to achieve a total surface rms of $\sim 35\,\mu$m, so as to allow operation right up to the atmospheric windows at 450 and 350 μm (~ 650 and $850\,$GHz). Using a terrestrial transmitter in the Fresnel region, working first at 31 GHz then 94 GHz, we found that the Misell algorithm enabled us to reduce the surface errors down from approximately 130 μm (the state of the surface after an initial mechanically based measurement/adjustment round) to approx. 80 μm (see Fig. 1(a)). However, further cycles of measurement/adjustment beyond this point did not seem to lead to any decrease in the real error, indicating that although the Misell algorithm was converging numerically, the answers it was giving for the surface phase were wrong, and the overall process was not converging. Other applications of the

23

Misell algorithm (e.g. Baars & Morris, 1988, and Whybourne & Morris, p.c.) have yielded a similar final accuracy for the total telescope rms (i.e. 60 to 80 μm), but at least in the case of the JCMT it was clear from the construction and panel errors that much lower figures should be possible. We found that the crucial step, enabling us to proceed to much lower rms surface errors, was to include *large scale fitting* as part of the algorithm. This was first noticed in connection with errors in pointing offsets for the transmitter, which was at a range of 700 m, corresponding to 1% of D^2/λ, and thus very much in the near field. An error in the assumed position of the transmitter (Δx, Δy) leads to a phase gradient, but completing the square in the quadratic Fresnel region term we see that this can be simulated by a bodily shift through ($\Delta x R$, $\Delta y R$) of the (u, v) pattern. This ambiguity leads to a poor performance of the Misell algorithm for the Fresnel region if pointing errors are present, with tell-tale signs of ghost images of blockage features. It might be thought that the solution is to fit for linear phase gradients in the aperture plane as the phase there is iteratively built up. This however still allows the solution to develop a component of translation, and in any case it can be shown that the basic Misell step is only in the 'correct' direction (in χ^2 gradient terms) if the phase error at any point of the aperture is $\lesssim 1$ radian (Lasenby, Waymont & Hills, in preparation). Thus in the presence of large phase errors due to a mispointing, or incorrectly known focus setting or even (in the HST case) some real intrinsic gross aberration of the telescope itself, the attempt to find these errors in the phase built up by the Misell algorithm itself in the aperture plane is unlikely to succeed, since it is precisely these phases that are in error due to the incorrect steps being taken by the Misell algorithm, in turn due to the presence of the large scale phase errors themselves. (Note these comments refer to runs with a limited number of Misell steps, say 50 maximum — there is some evidence (Anderson, p.c.) that longer runs begin to find such large scale terms, but this is very wasteful computationally.) Even if pointing errors (i.e large linear phase gradients) can be eliminated, we have found that the other sources of large scale error mentioned above, e.g. the wrong absolute focus position being assumed, can easily in practical circumstances lead to failure of the Misell algorithm (in the sense of incorrect answers below a certain rms accuracy, *not* due to noise). A further point is that fitting of large scale phase errors in the aperture plane (which we believe is what other workers generally do), can lead to difficulties with keeping track of 'wraparounds' through 2π as one moves around the plane. This is particularly so for the HST, where the spherical aberration leads to $\sim 8 \times 2\pi$ radians of phase change over the mirror surface.

Clearly, what is needed is a way of fitting the large scale phase effects directly in the beam plane, where none of these problems should arise. This was finally achieved by non-linear least squares fitting of low order Zernike polynomials, representing phase in the aperture plane, simultaneously to all the amplitude maps taken in the beam plane. The mathematical details are given in Lasenby, Waymont & Hills (in preparation), but the basic idea is to minimize the χ^2 difference between the measured amplitude points, and those that would be predicted for a given set of Zernike polynomial coefficients. The problem is computationally tractable, since all the necessary derivatives can be calculated by FFT's which are just weighted versions of the basic FFT which predicts the beam pattern given the aperture plane illumination. In the first version of this new algorithm, these 'large scale phase fits' were combined with Misell iterations for the smaller scale structure. However, it was soon realized that the same considerations which applied to the choice of a non-linear least squares algorithm working directly in the beam plane for the large scale fit, also applied to the small scale fit. Thus the algorithm now consists of both large-scale (i.e. low order Zernike polynomials) and small scale (i.e. pixel by pixel) least square fits, each of which works *simultaneously* with all the data measured in the beam plane. At present, a large scale fit is carried out every 20 small scale iterations, so as to 'relax' the results of the two processes together. Note by this means, any number of beam maps can be used, from one upwards, whereas the pure Misell algorithm only allowed two. This seems to work quite well, and for example, working typically with two defocussed 128×128 maps we have brought the *total* JCMT surface error down to $\sim 35\,\mu$m rms, with a component of measurement error of order 10 μm or less, since we are now reliably seeing structure in the individual *panels* of $\sim 10\,\mu$m amplitude. This structure, visible in the residuals shown in Fig. 2(a), is the 'surprise' mentioned above in connection with the JCMT. Simulations (see Fig. 2(b)) seem to show quite definitely that it is due to some combination of temperature induced warping of individual panels and a focal length misetting (resulting in a mismatch between the actual panel curvatures and those which would be needed for a dish with the current focal length) of ~ 16 mm.
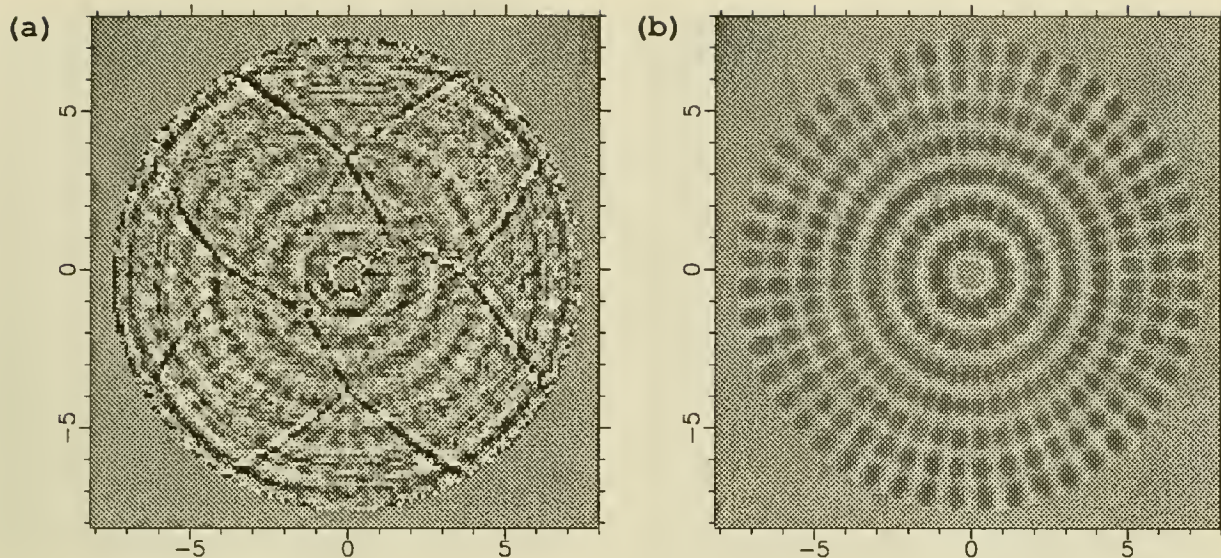
Figure 2: *(a) Residuals left in surface phase after fitting for the optimum movements of the panel adjustors. (b) A simulation showing the effect on surface phase of a 16 mm error in the focal length setting of the dish. The grey scales have different relative scaling, but the rms in each case is of order 10 μm.*

Having gone to an entirely least squares algorithm, means that one can now start viewing the process in 'likelihood' or 'Bayesian' terms. In particular, if it is assumed that the observed amplitudes are Gaussian distributed (thought to be roughly correct for the JCMT case), then the final phase and amplitude found for the aperture plane are the maximum likelihood estimators, or alternatively the Bayesian estimators if a uniform prior is assumed. This switch of view has three interesting consequences:

1. Regularizing functions can now be included easily if desired, on both amplitude and phase. If only one focus position is available for example, then the problem can be made better conditioned by including (e.g.) an entropy regularizing function of the form $S = \sum_i a_i \log(a_i/m_i)$, where a_i is the amplitude at the ith pixel of the aperture plane, and m_i is some default level (which can include information about the expected blockage) appropriate to this pixel. Writing the total function to be minimized as $\chi^2/2 - \alpha S$, we see that for Gaussian errors this is equivalent to maximising the likelihood, $\exp(-\chi^2/2)$, multiplied by the entropy prior $\exp(\alpha S)$ (see e.g. Skilling, 1989; Gull, 1989), so that the regularization can be put in a Bayesian context as well.

2. If the measurement errors are not in fact Gaussian, but have some other form, e.g. Poisson, then as long as we use the correct likelihood function for the data as measured, we can still carry out the process of maximising the product function 'likelihood × prior', where an entropy prior $\exp(\alpha S)$ can be used if conditioning is required, or a uniform one otherwise.

3. In both the above cases, proper account can be taken of measurement errors, missing points, etc., and confidence intervals found for the final result, i.e. we get a map of the surface plus errors.

(2) and (3) have not been fully implemented in practice yet, but both are conceptually straightforward within the current framework. Note that the actual algorithm used for the small-scale fit minimization, which in the HST case is with respect to $\sim 73,000$ independent parameters (a 192×192 complex grid), is of a Newton-Raphson type but needing only the main diagonal of the Hessian matrix, and seems to perform much faster than e.g. conjugate gradients, or other gradient search methods tried. All the work reported here can be carried out on a Vaxstation 3100 in about 4 hours maximum, although the memory requirements are large.
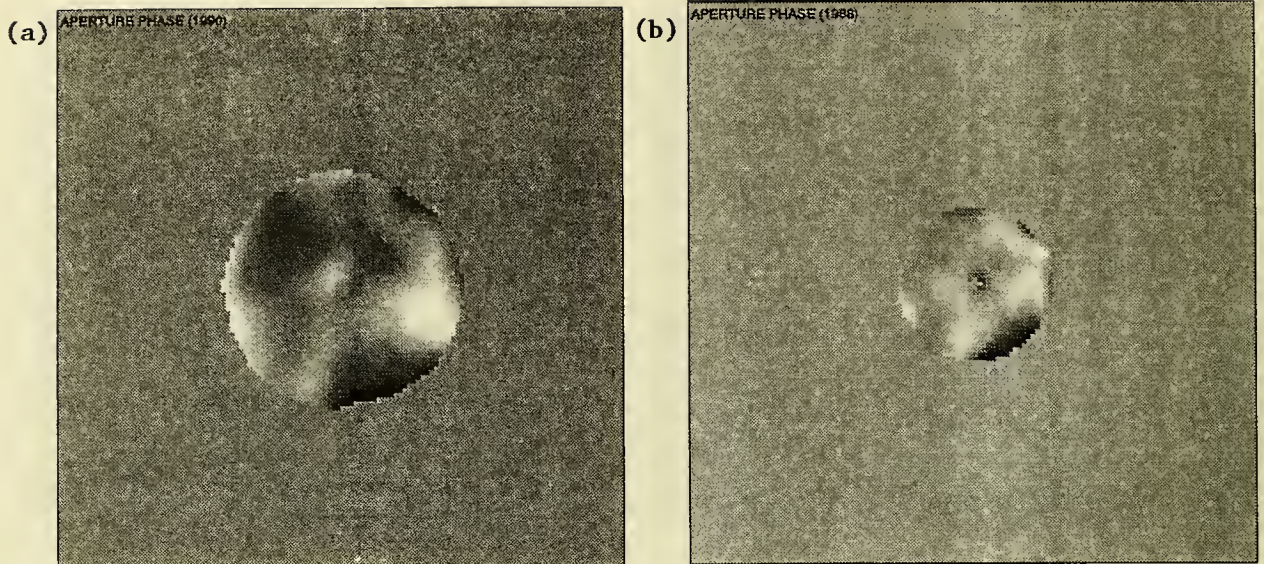
Figure 3: *Comparison of bolometer and heterodyne results for the surface of the JCMT, using Mars as a source. (a) The surface phase found in 1990 using a bolometer system working at 850 μm. (b) The surface found in 1988 using a spectral line receiver at 1300 μm. The rms of the large scale structure found is about 20 μm in each case and the different apparent dish sizes are due to different wavelengths and sampling intervals.*

3.1 Use of astronomical sources

Before discussing the results using HST star images, we first briefly demonstrate the method using astronomical sources and receivers, instead of a terrestrial transmitter and dedicated receiver, on the JCMT. The only sources strong enough at mm/sub-mm wavelengths are the planets, which have a non-negligible disc size, which must be taken account of in the program. The weaker signal strength compared to a terrestrial transmitter means that only lower resolution maps can be made, but being able to work at the high frequencies where the dish actually operates gives a corresponding increase in the sensitivity to deformations. Figs. 3, (a) and (b) shows a comparison of maps of the JCMT surface made using Mars as a source, on two occasions, 18 months apart. The point of the comparison is that despite the quite different receiver systems (one heterodyne and one bolometer) and wavelengths used, the results are in remarkable agreement, and indicate that we are detecting real structure on the surface of the dish, due in part to large scale features not so far taken out with the terrestrial holography, and in part due to real deformations of the telescope occurring as the dish is tilted up in elevation (the terrestrial holography takes place at an elevation of 11°, whereas both Mars maps were taken at ~ 50°). Note that given suitable sources, such maps can be made quite rapidly, without disturbing the astronomical configuration of the telescope at all, thus providing a useful diagnostic tool, which could be useful on other telescopes.

4 Preliminary results for the HST

Using out-of-focus stellar images on the HST is similar to a blend of the terrestrial transmitter and astronomical source cases on the JCMT. The signal is strong, with the necessary exposure times relatively short, and yet one is using the telescope directly at the wavelengths of interest, using real astronomical sources. As a further advantage, since CCD's are used, there is no necessity for raster scanning the telescope — unlike the JCMT, all of the beam is immediately available.

The data which will be discussed here are 5 Planetary Camera images, with 0″.043 pixel size, taken at 5470Å, with a filter bandwidth of about 500Å(F550M). This relatively large bandwidth is one of the
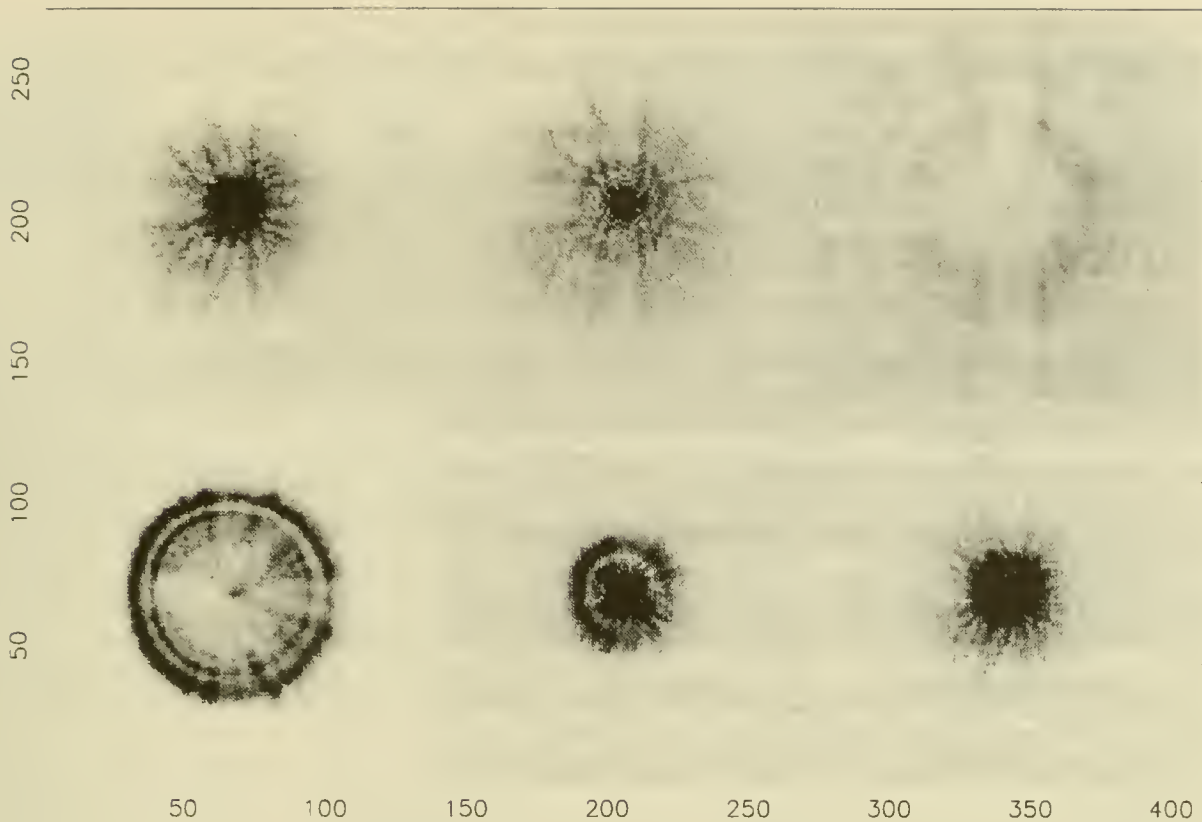
Figure 4: *The set of out-of-focus PC star images used for the phase retrieval results discussed in the text. From left to right on the bottom row the secondary mirror focus offsets are +333, +172 and +110 μm, while on the top row they are +49, 0 and −300 μm. The large asymmetry between the postive and negtive focus offsets immediately shows that some large intrinsic aberration is present.*

principal causes for concern with regard to the accuracy of the results presented below, since all the calculations assume a *monochromatic* filter. This will be discussed further below. The images are shown (in high contrast form) in Fig. 4. The sixth image, at a focal offset of −300 μm could not be used since as supplied it was truncated at the edges of the 137 × 137 grid. Each of the images was placed on a grid of 192 × 192 total size, so as to give a total extent in the aperture plane of 2.62 m — enough to accommodate the 2.4 m primary. The results of the large and small scale fits are shown in Fig. 5 and Table 1. For the large scale fits, Zernike polynomials up to order 21 are included, and the coefficients of some of the more interesting of these are shown in Table 1. It should be stressed that these polynomials are fitted simultaneously (i.e. no assumption of orthogonality is made — which would be incorrect anyway due to the large blockage) and that all the 5 beam maps are used *simultaneously*, both for the large and small scale fits. Fig 5 shows the actual beams as compared to the beams predicted with large and small scale terms included. The major differences seem to be a lack of radial smoothness as compared to the actual beams, which is due to the fact that the program is assuming a monochromatic image, rather than the 10% spread of frequencies in the real images. The amplitude of total spherical aberration found, ($\sim 4.2\,\mu$m of path length at the mirror edge), agrees well with estimates by Chris Burrows of STScI, but because of the problems with wide filter bandwidth and variable internal obscuration in the PC (see below), no claim is made here for the absolute reality of the split into r^4 and r^6 terms shown in Table 1, or for the coma terms, although the levels are quite plausible.

Fig. 6 shows the amplitude and phase derived from the small scale fit (i.e. all large scale terms have been subtracted off). For this preliminary attempt, the amplitude has been heavily constrained (via the entropy prior discussed above) to a model of the pupil blockage taken from OTA documents. On the other hand, the phase is unconstrained and shows three features which may be real (all at levels $\lesssim \lambda/20$ in terms of surface deformation):

```
+333 micron focus: (x,y) offset (1/100th arcsec):  1.97  8.86; Scale factor 9.34
+172 micron focus: (x,y) offset (1/100th arcsec):  2.16  6.38; Scale factor 8.56
+110 micron focus: (x,y) offset (1/100th arcsec): -1.03 -0.76; Scale factor 9.38
 +49 micron focus: (x,y) offset (1/100th arcsec):  2.04 -1.41; Scale factor 9.13
   0 micron focus: (x,y) offset (1/100th arcsec):  0.10  4.02; Scale factor 9.28


Global defocus:     106.0 microns
Global r**4 term:     3.45 microns
Global r**6 term:     0.61 microns
Global coma terms:    0.32 r**3 cos theta microns
                      0.20 r**3 sin theta microns
```

Table 1: *Actual results from the program for large scale phase fitting. Note that the total third and fifth order spherical aberration is the sum of the r^4 and r^6 terms, and that all results are quoted in terms of path length error at the mirror edge. If an r^6 term is not included the r^4 term is $4.2\,\mu m$ with a larger defocus. Note also that each map is allowed to have a separate multiplicative scale factor, which is fitted for.*

1. An annular ring, or possibly two rings with a gap between, stretching all the way round. Structures of this sort might well be left at the end of the grinding process.

2. An area in the inner SW quadrant where the process has failed to converge, and the interior zero grey level can be seen to extend out into the mirror proper. This may be real in the sense that it is known that because of the gross spherical aberration, internal obscuration in the PC is larger than it would otherwise be, and *varies* as a function of field position and (most unfortunately) focus position. This gives rise to the highly variable psf of the WFPC. In the present attempt at finding the mirror surface, no allowance has been made for this, and it is most likely that the region of non-convergence here is due to an obscuration which is present in some of the 5 images, but not all.

3. Similar remarks apply to the faint ghost images of the 'spider' legs, seen particularly to the right of the top and bottom legs. These multiple images could again be caused by the variable obscuration in the PC.

It should be stressed again that these results are highly preliminary, and based on only 5 out-of-focus images as against the large number which are potentially available. However, it is felt that it does represent an encouraging start, and that via this technique a great deal could be learnt about the internal optics of the HST.

5   Prospects for future work

The four main areas still to be explored are:

(i) Adapting the program to work with Faint Object Camera (FOC) images. These are noisier, but the ability to combine several of them, together with the lack of internal obscuration inside the FOC (at least in the f96 mode) may more than compensate for this.
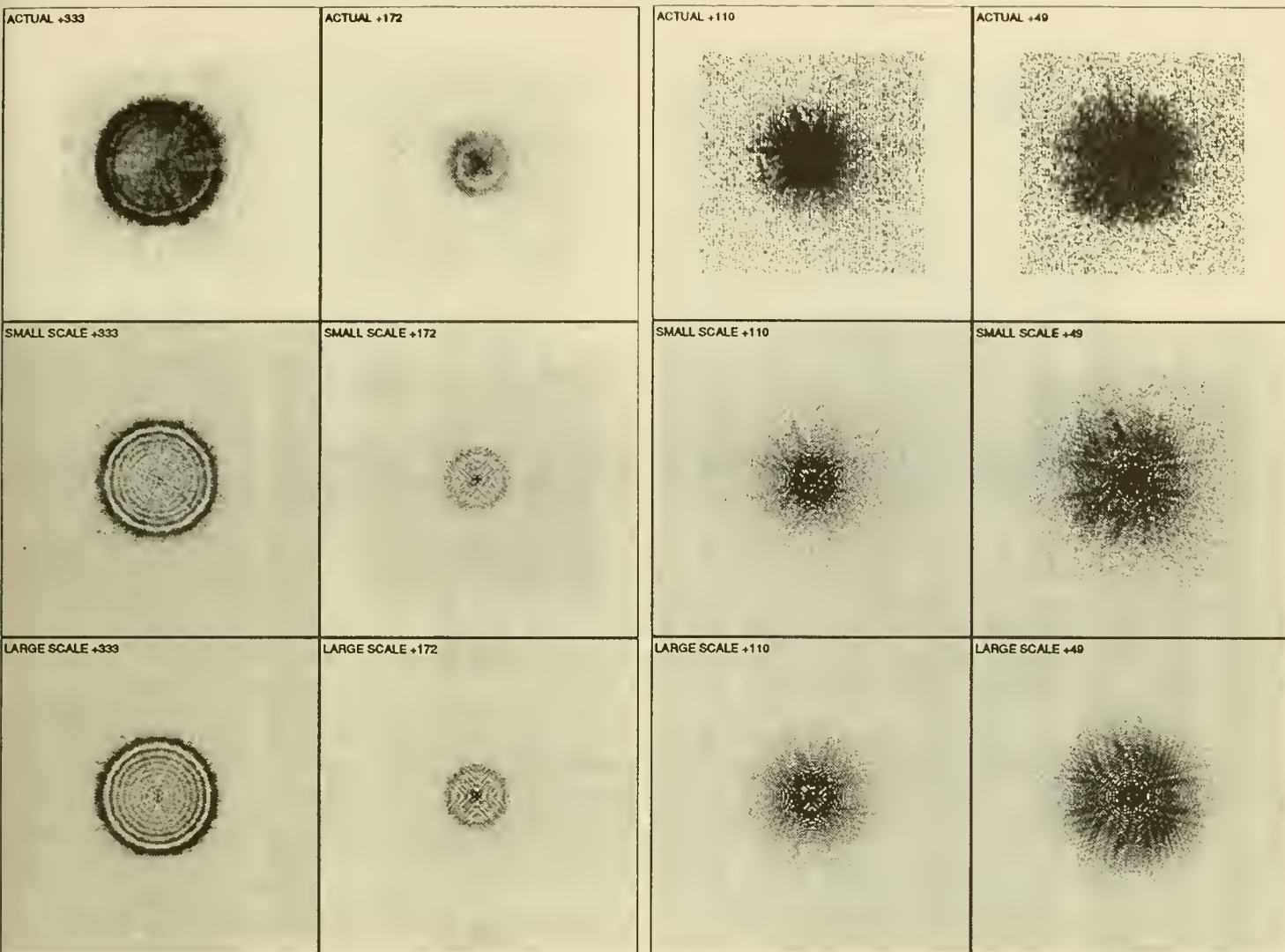
Figure 5: *Fitted and actual out-of-focus maps. In each column the top map shows the actual measured beam; the second row the result of both large and small scale fits and the bottom row the result with a fit for large scale terms only. The four columns are for the +333, +172, +110 and +49 μm offset cases respectively. (The 0 μm offset map was included in the calculations but not shown here.) In the last two columns the maps are shown at high contrast to highlight the extended features.*

(ii) Use of lower bandwidth filters, or alternatively extending the program to work properly with the finite bandwidth case. Note that simulating a psf with a wide bandwidth filter is easy — it is the inverse process which is more difficult.

(iii) Changing the least squares process to work with Poisson noise. This is particularly important for the FOC, where the low saturation levels mean that Poisson noise dominates. A shift to a full-scale likelihood approach will enable errors to be calculated properly, on both the large scale and small scale fits.

(iv) Developing a proper model for the variable obscuration inside the PC, and also using the data themselves to tell us features of the obscuration which may not be available from pre-flight drawings. The main initial step involved in the latter is to remove the maximum entropy constraint on the amplitude and let the program solve for a separate aperture plane amplitude for each focus setting (though a common phase).

Work on all of the above is already underway, and it is hoped to make substantial progress towards a much more accurate characterization of the HST optics over the next few months.
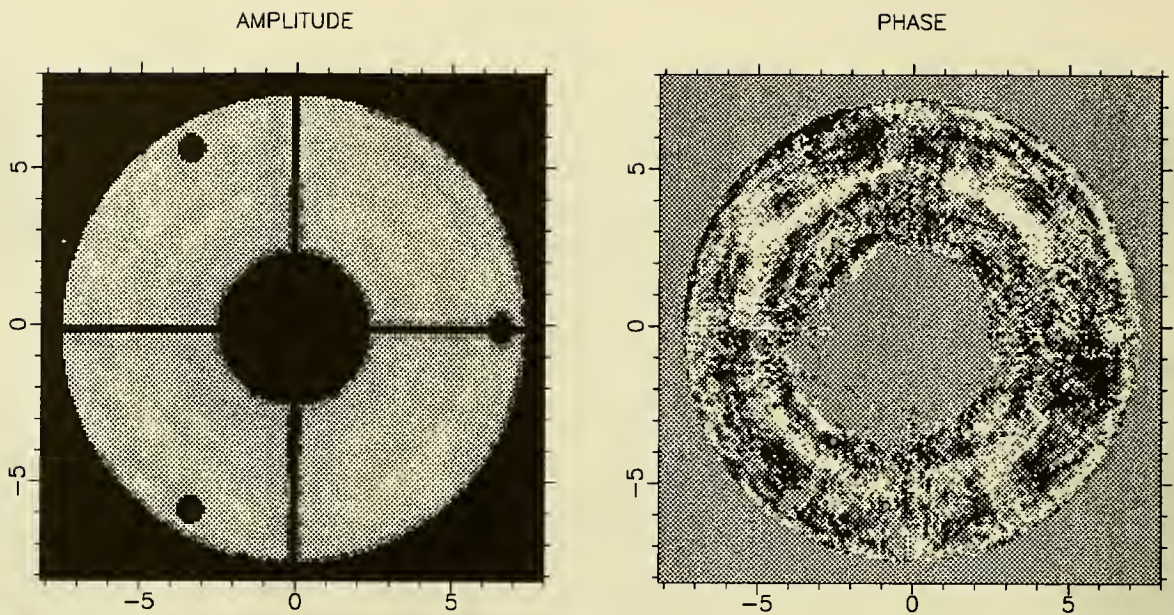
29

AMPLITUDE　　　　　　　　　　　　　　PHASE

Figure 6: *Reconstructed amplitude and phase for the HST, using 5 out-of-focus images. A heavy maximum entropy constraint has been placed on the amplitude, but no regularization was used for the phase. The largest phase deviations correspond to features on the mirror surface of about 25 nanometres. Particular points to note are discussed in the text.*

6　Acknowledgements

The work on the JCMT has been carried out in collaboration with Richard Hills, Devinder Sivia and Dave Waymont from MRAO, and invaluable assistance in the HST work was provided by Bob Thompson from the Institute of Astronomy, Cambridge. I would also like to thank Craig Mackay, Gerry Gilmore and Jim Fienup for useful conversations, and Ron Allen and others at STScI for encouragement and help.

7　References

Baars, J.W.M. & Morris, D., 1988. In 'Proceedings of the 11th ESTEC Antenna Workshop on Antenna Measurements', p43, ESA WPP Series.

Gull, S.F., 1989. In 'Maximum Entropy and Bayesian Methods', (Cambridge 1988), p53, ed. Skilling, J., Kluwer.

Lasenby, A.N., 1985. In *Proc.* 'ESO-Iram-Onsala Workshop on (sub)mm Astronomy', (Aspenas 1985), eds. Shaver, P. & Kjar, K., p77, ESO Conference & Workshop Proceedings, No. 22.

Morris, D., 1985. *IEEE Trans.Ant.Propag.*, **AP33**, 749.

Padin, S., Davis, R.J. & Lasenby, A.N., 1987. *Mon.Not.R.astr.Soc.*, **224**, 685.

Skilling, J., 1989. In 'Maximum Entropy and Bayesian Methods', (Cambridge 1988), p45, ed. Skilling, J., Kluwer.

MEM: NEW TECHNIQUES, APPLICATIONS, AND PHOTOMETRY

N. Weir and S. Djorgovski
Palomar Observatory
Division of Physics, Mathematics, and Astronomy
California Institute of Technology, 105-24
Pasadena, California 91125 USA

ABSTRACT. We examine a new implementation of the Maximum Entropy (MaxEnt) method of image restoration for direct astronomical images. Recent advances in the theory and application of MaxEnt have led to a significantly improved version of the algorithm, rendering virtually all previous MaxEnt codes obsolete. The use of a subpixelated image space, intrinsic correlation functions, and an objective, theoretically justified stopping criterion combine to lend the new method increased power. We demonstrate the results of the algorithm applied to simulated and real data. We find that it is possible to recover many features at a resolution approaching the pixel size in the original data, even when such features are embedded in extended structure. We compare the results of using this new version of MaxEnt, an older version, and the Richardson-Lucy algorithm on an actual HST WFC image of R136. The photometry of sources in the restored images are systematically biased downward by about one sigma, as expected. We suggest how the MaxEnt result may be used as an initial input to a standard least squares model fitting program to achieve unbiased photometry on objects detected in the deconvolved image. We demonstrate a first attempt at this hybrid approach on a simulated HST WFC image of a dense star cluster.

1. Introduction

Maximum Entropy (MaxEnt) deconvolution algorithms have been implemented and applied in various forms for well over a decade. Not since the introduction of MaxEnt to the image restoration problem, however, has there been as fundamental an advancement in the power of the technique as has taken place in the last two years. While the result is images of a resolution and accuracy not known to be achievable as recently as a year ago, there are still aspects of the restoration process which require careful analysis for proper interpretation of the results (e.g., photometric bias). We examine this latest method, paying particular attention to its morphological and photometric reliability.

The restoration software used in this study is an implementation of the MEMSYS-3 package (Gull and Skilling, 1989). A general description of the new MEMSYS algorithm and the significant improvements of this compared to all previous versions of their code may be found in Skilling (1990) and Gull (1990). A driver program adapting their routines to direct astronomical imaging applications was developed at Caltech and will be described in more detail elsewhere (Weir, 1991). The most significant aspects of the new, "Quantified" MEMSYS-3 are that it objectively determines the appropriate stopping criterion fully within a Bayesian framework, and it allows for the estimation of the flux and an error bar for any portion of a restored image (cf. Skilling, these proceedings). The novel aspects of the Caltech implementation are described in the next section.

2. New Techniques

A little appreciated, but quite significant, aspect to the general image restoration problem is that the data are defined in data space and restored images are defined in image space, and the two spaces *need not be identical*. In other words, there is no *a priori* reason that the pixel grid within which the image is to be restored should be exactly the same scale as the data. In fact, to extract the full information content of the data, it is generally necessary to restore to a more finely sampled grid.
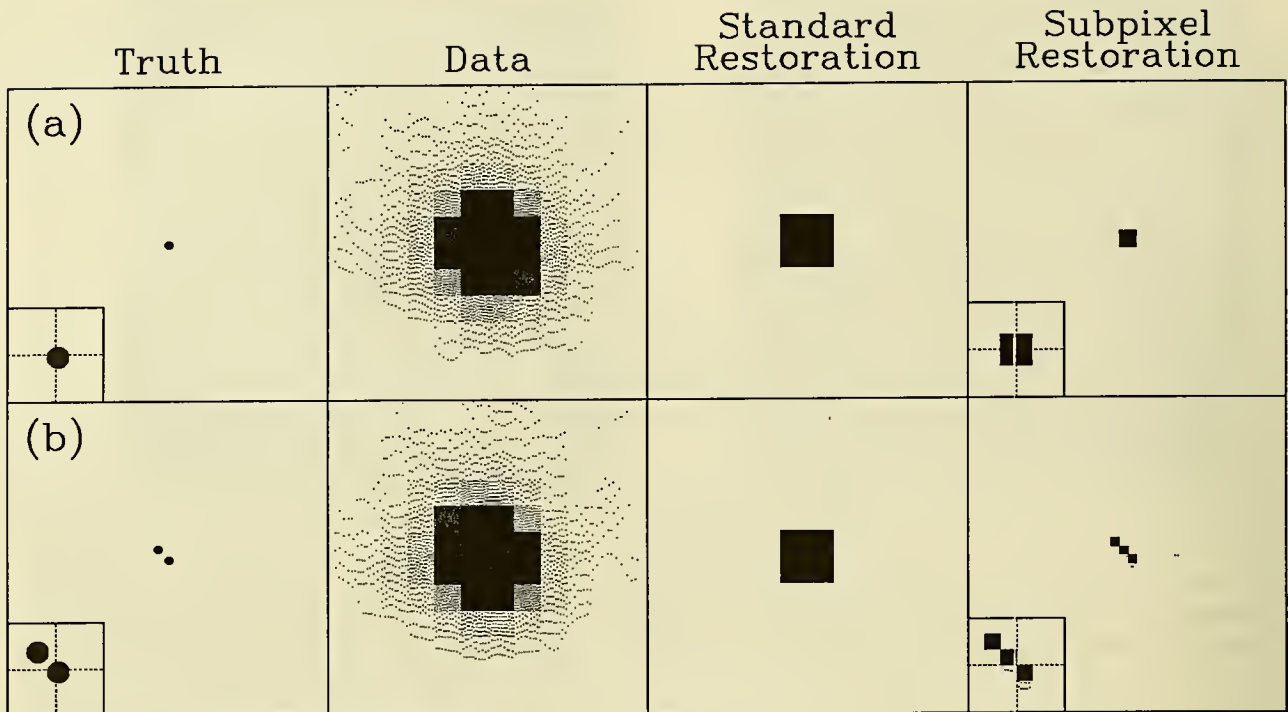
Fig. 1.   Simulated images and restorations of a point source (a) and a binary (b) of separation 0.6 pixel. The insets depict the subpixel location of the original sources and their subpixel restoration within a grid of two-squared data pixels. The second restoration was performed using three times higher spatial sampling than in the first, resulting in nine image pixels for every data pixel.

Real information exists at these scales, and we commonly use it when we find centroids of objects to subpixel accuracy. By forming a subpixelated image, one is simply creating an enlarged hypothesis space within which to obtain a high degree of positional precision. As will be demonstrated below, this technique also facilitates the detection of very high resolution structure in the restored image which otherwise might not be apparent due to the large pixel size of the data. From simulated images and double blind tests, we have never found our restoration method to introduce structure at subpixel scales when it did not actually exist (Weir and Djorgovski, 1991).

To restore to such levels, one must be able to adequately interpolate the point spread function (PSF) at the subpixel level. We typically use a PSF determined by the stellar photometry program Daophot (Stetson 1987), which achieves a three times higher than nominal sampling estimate of the PSF by forming a composite of stars from the image of interest. A current image estimate is convolved with this finely sampled PSF and subsequently rebinned in order to compare against the data. The data in their original form, therefore, are always used for assessing the goodness of fit of the present image. Some algorithms demand that the data themselves be expanded or interpolated in order to restore to a more finely sampled image. Such transformations of the data, however, render their noise properties ill-defined.

Figure 1 illustrates a case where detailed structure is readily apparent only when restored with subpixelation. Figure 1a depicts a point source, while 1b contains a binary of equal intensity and separation 0.6 pixels. The combined intensity of the binary is the same as for the single source, and an equal background and amount of signal dependent and independent noise have been added to both. The SNR of the peak pixel in each image is just under 140. The point sources were centered within individual pixels at the locations illustrated in the insets of the first panel of 1a and 1b. The PSF used to create these images was empirically derived from an actual ground-based CCD image and has a full width at half maximum (FWHM) of approximately 2.25 pixels. Note the power and accuracy of the subpixel restoration result.
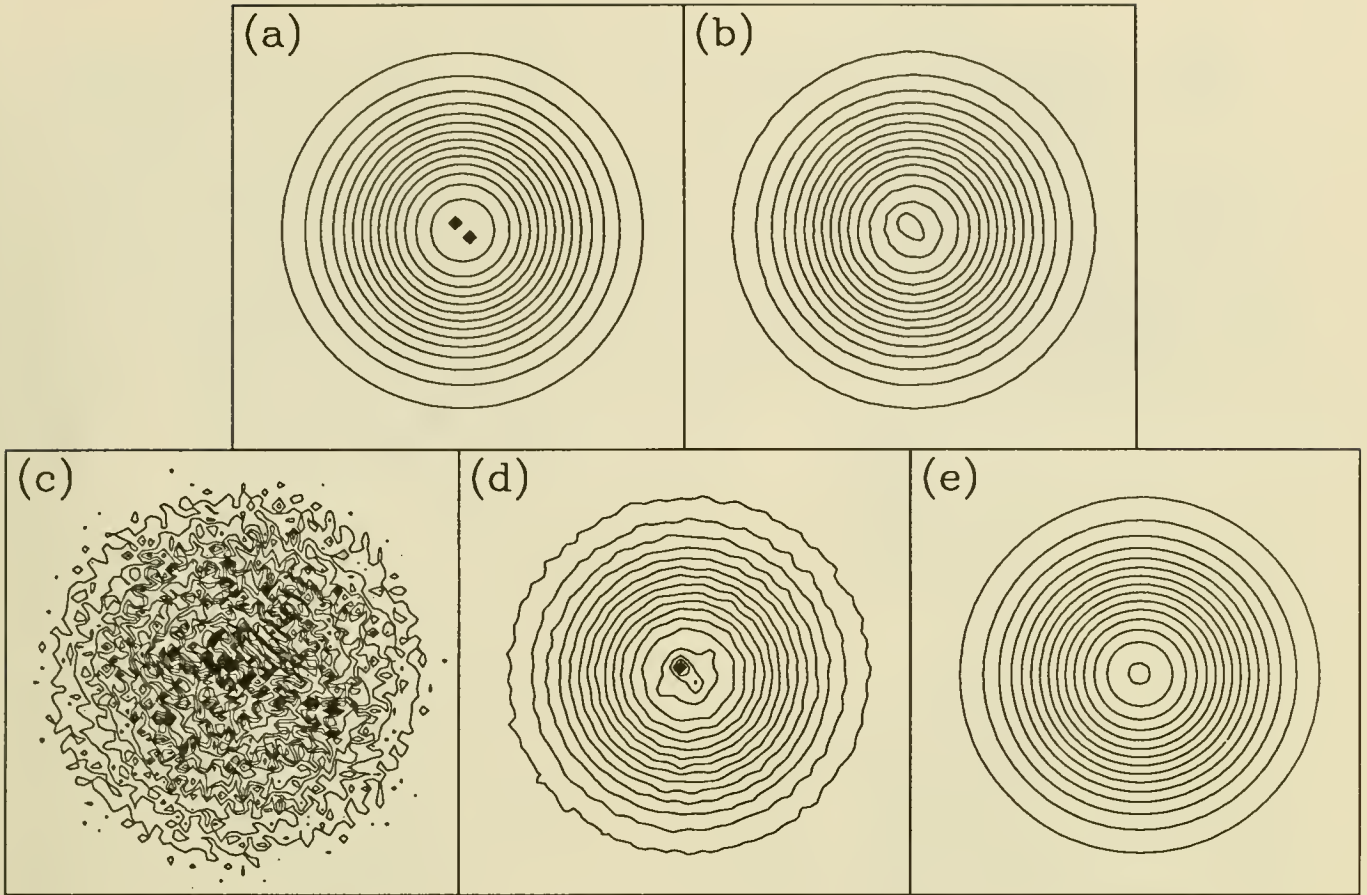
Fig. 2. Restoration of two point sources of intensity ratio 2:1 lying atop an extended background. (*a*) The "truth." (*b*) Simulated data obtained by convolving (*a*) with a Gaussian PSF of $FWHM = 3$ pixels and adding signal dependent and independent noise. Restorations were obtained using (*c*) no ICF, (*d*) two Gaussian ICFs where $\sigma_1 = 0.6$ pixel and $\sigma_2 = 5$ pixels, and (*e*) one ICF with $\sigma = 5$ pixels. Contours are spaced logarithmically in intensity and are equal for all five images.

A second major advance in our latest implementation of MaxEnt is in the use of intrinsic correlation functions (ICFs). ICFs may be operationally defined as follows. Imagine that the restored image we create using MEMSYS is in fact a "hidden map." This image does not actually represent our final estimate of the true object distribution, but must first be convolved with some kernel (the ICF), which may be of Gaussian form, for example. The ICF introduces a degree of spatial correlation in the restored image. This convolved image, in turn, must be convolved with the PSF in order to create a "mock" data set which may be directly compared with the data. The practical benefit of the ICF is to give the final image a smoother, more aesthetically pleasing appearance. It may be rationalized in terms of probability theory as introducing extra prior information into the restoration method: the hypothesis space, or space of feasible images, is limited to those images with at least some degree of spatial correlation.

The choice of what degree of correlation to enforce is in most cases subjective. but not always critical. If one uses a broad ICF, it will dampen out high spatial frequency noise and ringing in the final image, but at the expense of washing out real point sources as well. In theory, one could try restoring using many different forms for the ICF, and the value of the "evidence" returned by MEMSYS 3, the log of the probability of the data given the model hypothesis, should be the largest for the ICF most well-suited to the data. In practice, it takes far too long to search the space of ICFs to find a maximum.

We have found the following technique to be the most powerful, albeit computationally expensive.
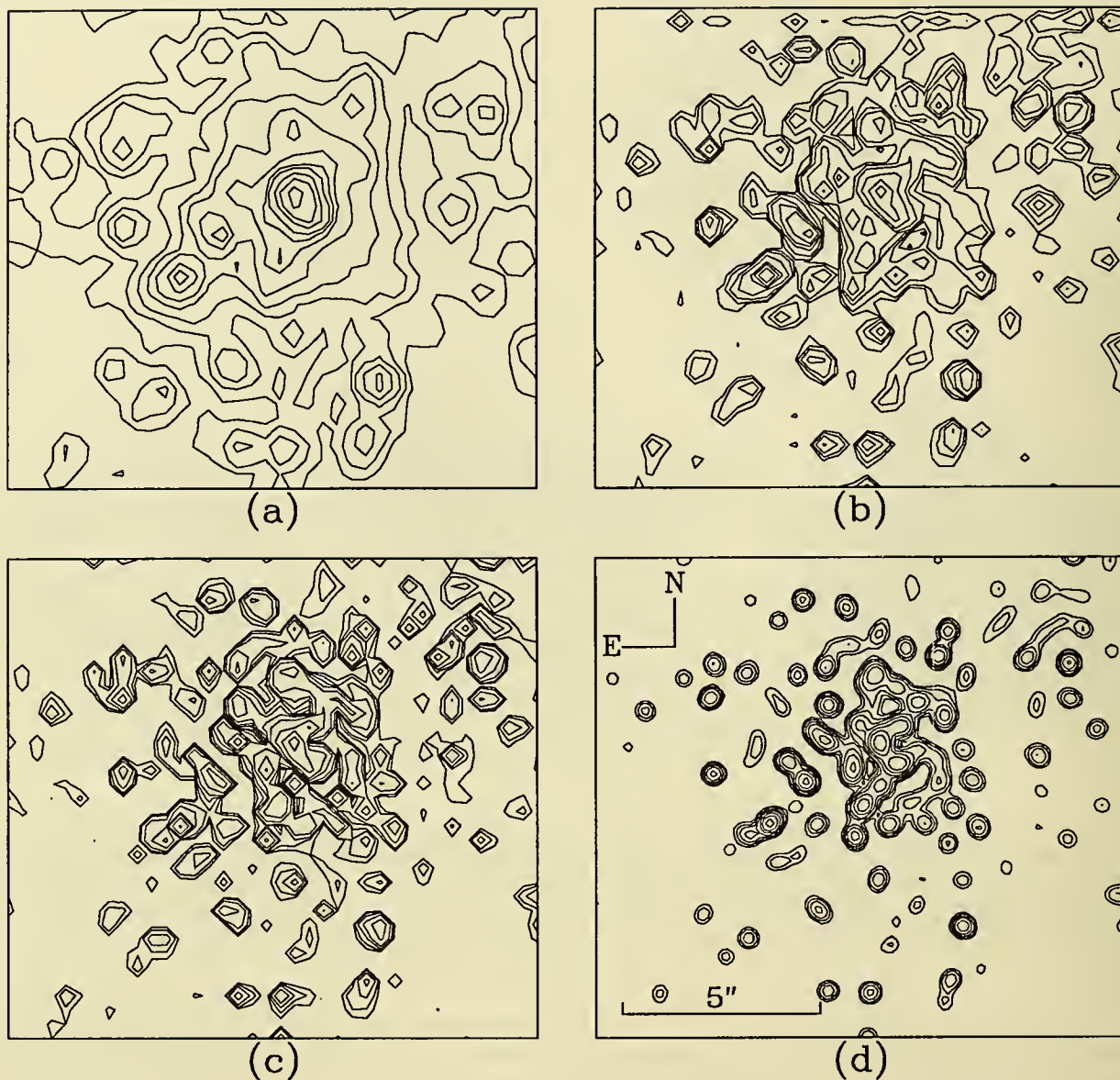
Fig. 3. Original image (*a*) and restorations (*b* − *d*) of the star cluster R136, from a V band CCD image obtained at European Southern Observatory (see Weir, Djorgovski, and Meylan, 1991). (*b*) R-L result after 60 iterations; (*c*) "standard" MaxEnt result; (*d*) latest MaxEnt result using subpixelation and an ICF. Contours are logarithmically spaced in intensity.

We simultaneously restore to two hidden maps, one which has a relatively broad ICF, the other having a very sharp ICF, or none at all. The final result is sum of the convolution of the two hidden maps with their respective ICFs. Using this approach, most of the extended structure is restored to the channel with the broad ICF, thus retaining its smooth appearance; the strong point sources and high spatial frequency structure (and a diminished amount of ringing and noise) is restored to the narrow ICF channel. Figure 2 illustrates the results of this technique applied to data containing two point sources lying on top of an extended Gaussian. Such a source is notoriously difficult to restore using standard deconvolution methods. Figure 2*c* shows the result of restoring using our MaxEnt code without an ICF. The method clearly over-resolves the data. Figure 2*e* demonstrates the result of
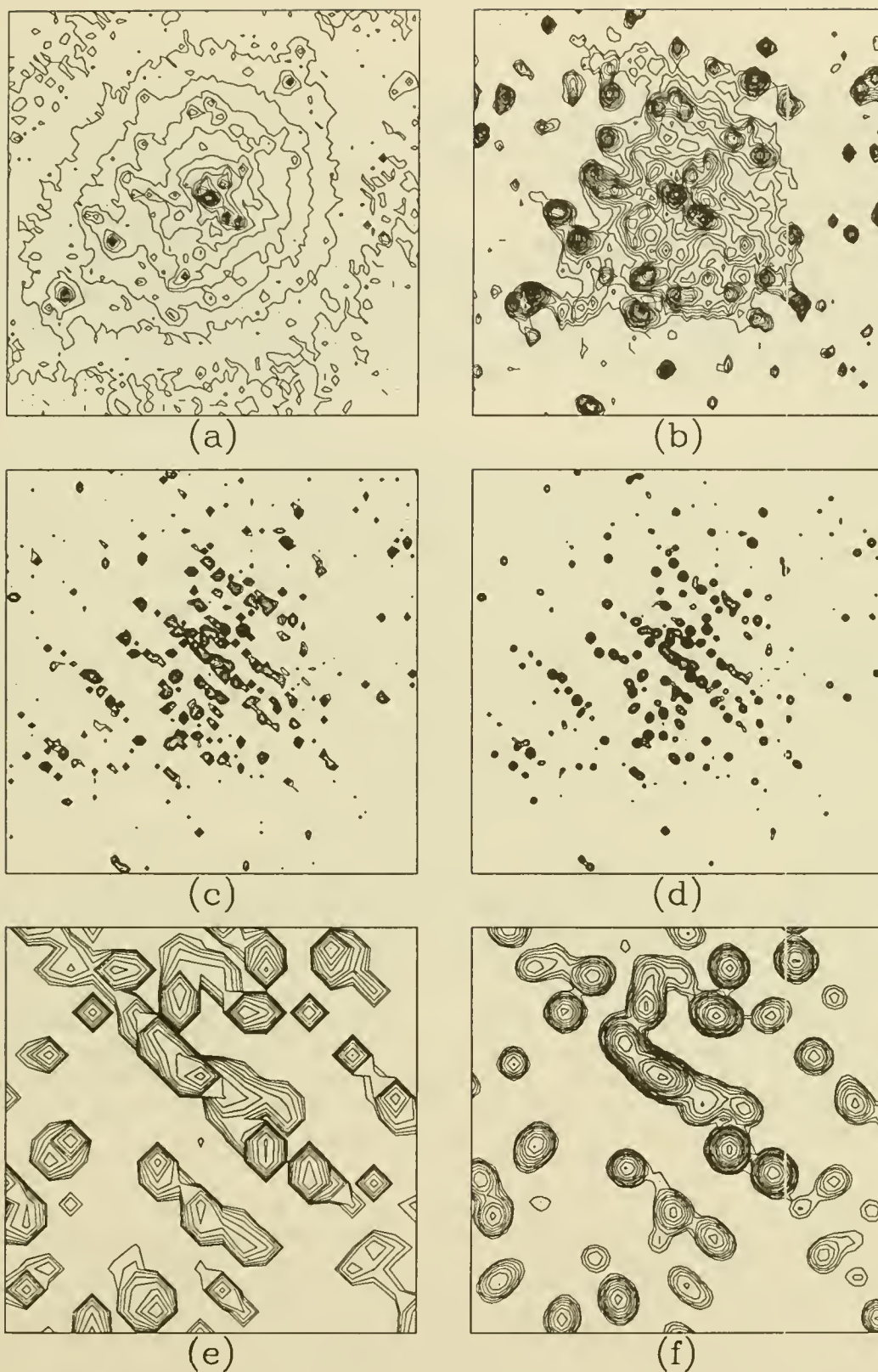
Fig. 4. Original HST WFC image (*a*) and restorations (*b* − *f*) of the star cluster R136 (see *STScI Newsletter*, August 1990, Vol. 7, No. 2). (*b*) R-L result after 30 iterations; (*c*) "standard" MaxEnt result; (*d*) latest MaxEnt result using subpixelation and an ICF. North is up, East to the left, and images are 8 arcecs on a side. Note the different scale and field size as compared to Figure 3. (*e*) and (*f*) are 4× expansions of the centers of the two MaxEnt restorations, (*c*) and (*d*), above them. Contours are logarithmically spaced in intensity.

using a broad Gaussian ICF of $\sigma = 5$ pixels. The extended background is very accurately recovered, at the expense of completely losing the two point sources. Figure 2d is the result of simultaneously restoring to two hidden maps with different ICFs: one with $\sigma = 0.6$ pixels, the other with $\sigma = 5$ pixels. Note that both of the point sources are readily detected in this image with only slight ringing in the background structure.

In general, we find that the final result is relatively insensitive to the initial flux ratio of the two hidden map components. Thus, the primary subjectivity in applying this method is in deciding what to use for the two (or more) ICFs. First indications are that as long as one ICF is a few times broader than the PSF and the other is narrower, the final results are quite similar. This approach works best when used in conjunction with subpixelation, thereby allowing for the possible recovery of detail at the smallest scales while still obtaining the benefits of spatial correlation.

3. Applications

In Figure 3 we present the results of applying three different restoration methods to a ground-based image of the stellar field, R136: the dense core of the 30 Doradus star cluster in the Large Magellanic Cloud. Our image of R136 was obtained in a single three second exposure and represents the best ground-based observation of the cluster obtained to date (Weir, Djorgovski, and Meylan, 1991). The FWHM of the PSF in the original image is about two pixels, corresponding to approximately 0.7 arcsec. We used the same PSF estimate derived from the data in all three methods. Figure 3b depicts the result of applying 60 Richardson-Lucy (R-L, from Heasley, 1984) iterations. This result compares quite reasonably with image 3c, the result of applying a "standard" MaxEnt deconvolution algorithm, using no subpixelation or ICF, and employing the traditional $\chi^2 = n$, the number of data points, stopping criterion. Figure 3d is the result of using the latest MaxEnt algorithm with subpixelation and a Gaussian ICF of $\sigma = 1$ restored image pixel. The latter restoration is not only the most easy to interpret, but it brings out structure not evident in either of the previous results. The improvement in resolution in the final restored image is about a factor of three.

The structure at the very center of the restored cluster core corresponds directly to a blurred version of the speckle result taken at the same wavelengths (Weigelt and Baier, 1985). Reliability at larger scales is confirmed by analyzing an image of the same cluster obtained by the HST WFC (see Space Telescope Science Institute Newsletter, August 1990, Vol. 7, No. 2). A portion of this image is presented as Figure 4a. Figures 4b, 4c, and 4d, represent the result of applying the R-L algorithm (30 iterations), "standard" MaxEnt, and new MaxEnt to these data. Figures 4e and 4f are 4× expansions of the very center of 4c and 4d, respectively. Again, the same PSF was utilized for all three restoration methods, and was derived empirically from the data using Daophot. Note the high level of correspondence between the ground-based MaxEnt result presented in Figure 3d and the raw and R-L restored HST images. Verifying the power and accuracy of that restoration, we are able to establish a degree of confidence in the MaxEnt deconvolutions of the HST data. The MaxEnt results appear to be vastly superior to those produced by R-L iterations.

4. Photometry

The photometry of sources in MaxEnt restorations is known to be systematically biased by approximately one sigma in the downward direction (e.g., see Sibisi, 1990, and Cohen, 1991). We have found that the bias can be modeled through Monte Carlo simulation, providing the possibility of statistically correcting for this effect in a given image. We, however, prefer the following approach. Given that MaxEnt does an excellent job of object detection and separation, why not use the restored image as a high-resolution "finding chart" by which to obtain first estimates of the position and flux of all objects in the image? One can then feed these estimates into a least squares model fitting package, such as Daophot, optimized for the actual HST PSF shapes (see, e.g., Holtzman, 1990), to obtain unbiased photometry from the original, lower resolution data.

To demonstrate this bias and how one might correct for it using a hybrid MaxEnt/model fitting approach, we created a simulated image of a star cluster containing 139 stars spanning a range of about five magnitudes and peak pixel S/N of ~ 150 to 10. We used the sample WFC PSF provided
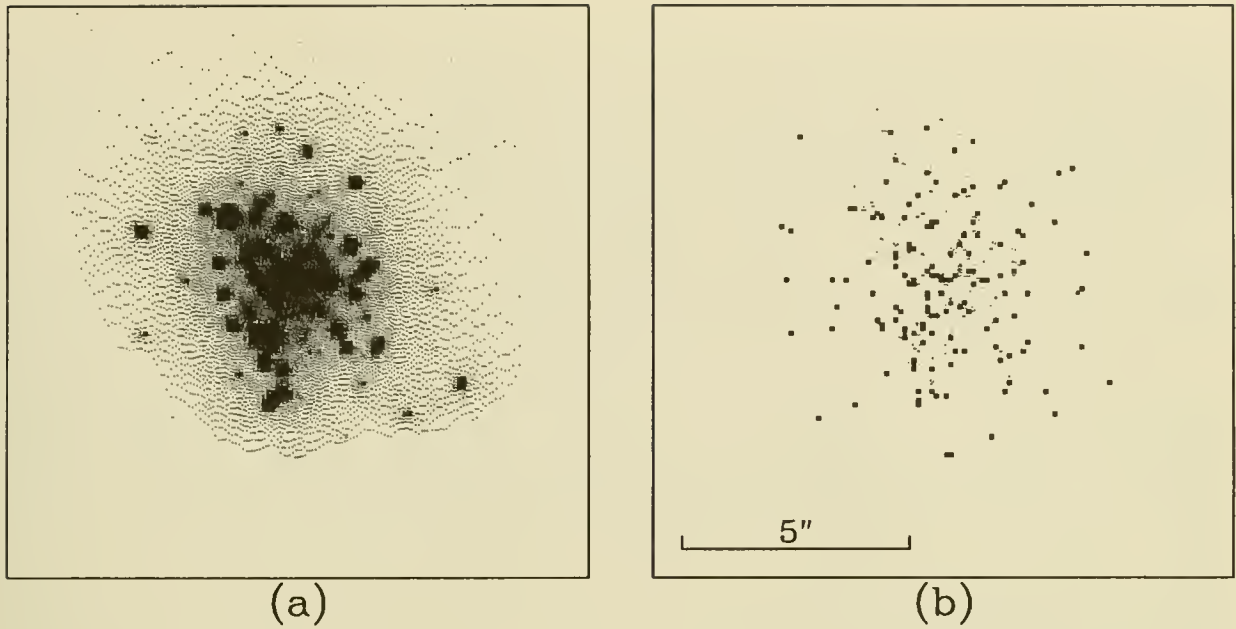
Fig. 5. (*a*) Simulated HST WFC image of a star cluster and (*b*) its restoration. The latest MaxEnt version has been used without ICFs or subpixelation. Stars were placed at the center of individual pixels in the simulated image so that they would restore to individual pixels in the final image.

Fig. 6. Photometric results for all real, detected stars in the original restored image, and in the revised image described in the text. The insets are histograms of the magnitude errors for all stars brighter than 16 mag in the corresponding restored images.

by the WFPC team for testing deconvolution methods on simulated HST data. To facilitate the easy detection and separation of the stars in the restored image, we placed all stars at the centers of specific pixels. As a result, all of the flux for each s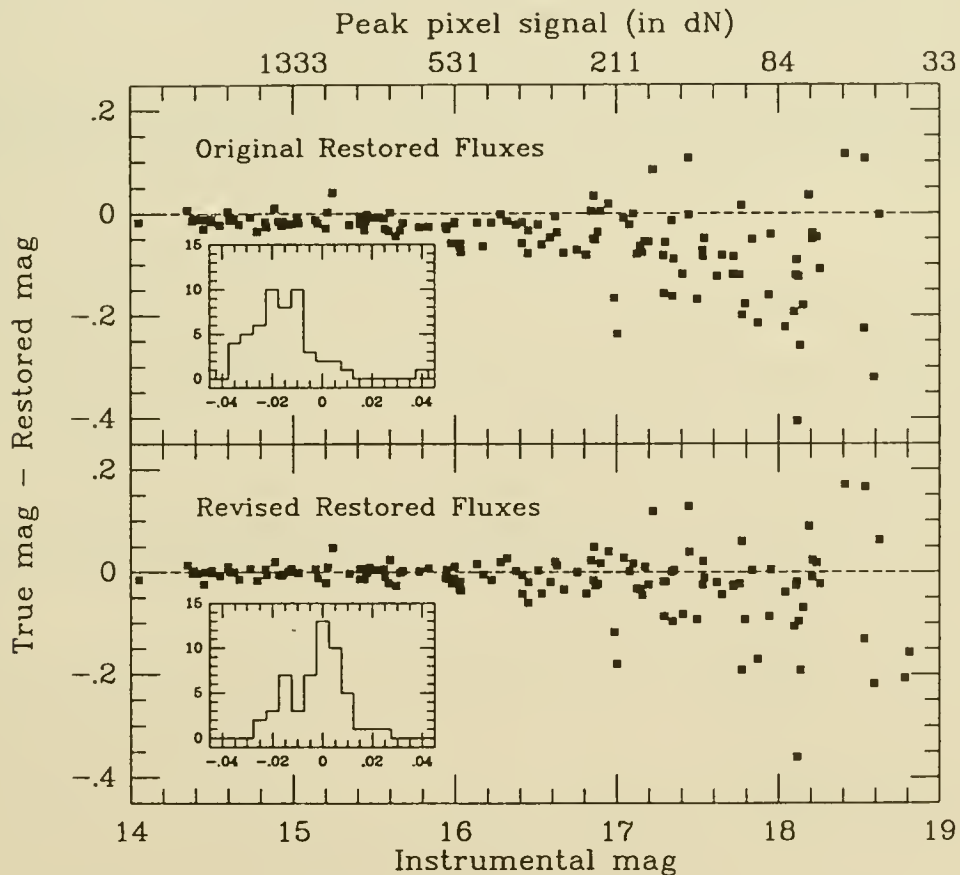tar restored to individual pixels within the final image, allowing us to perform photometry simply by associating stars with individual pixels over a given threshold in the image.

The result of restoring the simulated cluster using the latest MEMSYS algorithm, but without subpixelation or ICFs, is shown in Figure 5. All of the stars were found, with only one false detection within the restored magnitude range spanned by the real stars. The top panel of Figure 6 demonstrates how the magnitudes of stars in this "original" restored image are systematically biased downward.

The next step should be to feed all of the detected stars to a model fitting procedure. Instead, simply for demonstration purposes, we took the following shortcut suggested by John Skilling. We applied a threshold to the restored image, setting to zero all pixels below a certain value. All pixels above the threshold were assumed to represent real stars. We then adjusted a scaling factor within MEMSYS 3 so that in the function to be maximized at each iteration, the image entropy term weighed negligibly compared to the likelihood function, which measured the misfit between the data and the current restoration. Thereafter, the code served essentially as a maximum likelihood model fitting program, adjusting all of the non-zero image values to match the data in a least squares sense. The resulting "revised" restored fluxes are plotted in the lower panel of Figure 6. Note that while there still exists some bias at the faintest end, the errors in the restored magnitudes are distributed in a much more nearly Gaussian fashion about zero. A more accurate and sophisticated model fitting procedure will clearly be needed to extend this approach to real data where, for example, the one pixel/one star assumption certainly breaks down. Nonetheless, these results indicate that we may be proceeding on the right track in extending the scientific usefulness of MaxEnt deconvolutions beyond the realm of morphological analysis, to improved imaging photometry.

REFERENCES

Cohen, J. (1991) in W.T. Grandy, Jr. and L. Schick (eds.), *Maximum Entropy and Bayesian Methods*, Kluwer, Dordrecht, in press.

Gull, S.F. (1990) in J. Skilling (ed.), *Maximum Entropy and Bayesian Methods*, Kluwer, Dordrecht.

Gull, S.F. and Skilling, J. (1989) *Quantified Maximum Entropy "MEMSYS 3" User's Manual.*

Heasley, J.N. (1984) *Publ. Astron. Soc. Pac.* **96**, 767.

Holtzman, J.A. (1990) *Publ. Astron. Soc. Pac.* **102**, 806.

Sibisi, S. (1990) in J. Skilling (ed.), *Maximum Entropy and Bayesian Methods*, Kluwer, Dordrecht.

Skilling, J. (1990) in J. Skilling (ed.), *Maximum Entropy and Bayesian Methods*, Kluwer, Dordrecht.

Stetson, P. (1987) *Publ. Astron. Soc. Pac.* **99**, 191.

Weigelt, G. and Baier, G. (1985) *Astron. Astrophys.* **150**, L18.

Weir, N., Djorgovski, S., and Meylan, G. (1991) in K. Janes (ed.), *Formation and Evolution of Star Clusters, P.A.S.P. Conf. Ser.*, in press.

Weir, N., and Djorgovski (1991) in W.T. Grandy, Jr. and L. Schick (eds.), *Maximum Entropy and Bayesian Methods*, Kluwer, Dordrecht, in press.

Weir, N. (1991) in preparation.

BLOCK ITERATIVE RESTORATION OF ASTRONOMICAL IMAGES FROM THE HUBBLE SPACE TELESCOPE

Don J. Lindler

Advanced Computer Concepts
Potomac, Maryland

1. INTRODUCTION

The discrete model of linear image degradation is specified by the equation:

$$\mathbf{b} = H\mathbf{x} - \mathbf{n} \tag{1}$$

where **b** and **x** are the pixel values of the degraded and original undegraded images stacked into column vectors, H is a matrix constructed from the impulse response (or point spread function) of the degradation, and n is an unknown additive noise vector. The object of restoration is to determine **x**, given **b** and possibly information on the properties of **n**. If the point spread function used to construct H is not known for the given optical- detector configuration, it must be estimated from the blurred image, **b**. The point spread function is most easily estimated from point sources (*i.e.* stars) on the blurred image.

Since H may be ill-conditioned or singular, and only the statistical properties of the noise are known, there are many solutions for **x** which satisfy equation (1). The success of a restoration therefore depends on the ability to model and apply to the restoration, known or assumed properties of the desired solution, such as positivity or smoothness.

Some advantages of algebraic image restoration are:

1) The point spread function may be spatially variant;

2) If a constrained least squares method is used, the applied constraints may be varied from pixel to pixel to make maximum use of the known image properties;

3) Missing or bad pixel values in the blurred images can be easily handled without attempting to repair their values;

4) Noise properties can vary from pixel to pixel.

The main disadvantage of algebraic image restoration is the size of the linear system. For a 500×500 pixel image, H is a 250.000×250.000 matrix. Even with the most powerful computers available, a direct solution of the system would be impossible. In the next section, we describe a technique – the block iterative method, of solving large linear systems.

2. THE BLOCK ITERATIVE RESTORATION ALGORITHM

2.1 Block Jacobi Iteration

In most astronomical images, the point spread function has a much smaller spatial extent than the image, so it is appropriate to work on the image locally. We therefore divide the image into blocks and restore each block separately, using values from the previous iteration as estimates of the unblurred image values outside the block. In most instances the blurred image

is a good choice for the starting or zeroth iteration. This type of iteration is called block Jacobi or group Jacobi iteration (Young 1971) and can be formulated in matrix notation as follows.

Consider the blurred image, **b**, divided into m blocks of equal size B_i, i = 1,m:

$$B = \begin{pmatrix} B_1 & B_2 & \cdots & \cdots \\ & B_{i-1} & B_i & B_{i+1} \\ & \cdots & & \\ & & B_{m-1} & B_m \end{pmatrix}$$

Stack the elements of each block and place them into a vector:

$$\mathbf{B} = (B_1, B_2, \cdots, B_m)^T$$

Ignoring the noise for now, we write the system as:

$$H\mathbf{X} = \mathbf{B}$$

where H is partitioned into blocks

$$H = \begin{pmatrix} H_{11} & H_{12} & \cdots & H_{1m} \\ H_{21} & H_{22} & \cdots & H_{2m} \\ & & \cdots & \\ H_{m1} & H_{m2} & \cdots & H_{mm} \end{pmatrix}$$

and **X** contains the restored values, blocked in the same manner as **B**. If the image were divided into blocks of n pixels each, then the block H_{ij} would have size $n \times n$. The block Jacobi method can now be written as:

$$H_{ii}X_i^{r+1} = B_i - \sum_{j=1, j \neq i}^{m} H_{ij}X_j^r \tag{2}$$

$i = 1, \cdots, m$, and where X_i^r is the stacked values for iteration r of block j. If we define the vector on the right hand side of equation (2) as BMOD_i (*i.e.*, the blurred image less contributions from outside the block as estimated from the previous iteration), the linear system for block i can now be written as:

$$H_{ii}X^{r+1} = BMOD_i \tag{3}$$

Using the block Jacobi method, we can reduce the problem to solving

$$H\mathbf{x} = \mathbf{b} \tag{4}$$

where H is H_{ii} for block i; **x** is X_i^{r+1} for block i and iteration r; and **b** is BMOD_i for block i.

The solution for block i now requires the solution of an $n \times n$ linear system. For example, to restore a 100×100 pixel image divided into m=100 blocks, each of size $n \times n = 10 \times 10$, the largest system to be solved would have H_{ii} of size 100×100. Since solutions of linear systems require on the order of n^3 operations, the block approach compares favorably to the direct solution of the $10,000 \times 10,000$ system. For a spatially invariant point spread function, the problem is further reduced because H_{ii} will be identical for all $i = 1, \cdots, m$.

If a constrained least squares approach is used to solve the linear system, the solution will converge to acceptable results even with a block size as small as the full-width-at-half-maximum (FWHM) of the point spread function. Overlapping the blocks (accepting only the central portion for the next iteration) can be used to speed convergence.

2.2 Image Constraints

The block Jacobi method reduces the restoration to solution of many smaller linear systems, but it does not address the ill-conditioned nature of H or the presence of noise in the blurred image. An ill-conditioned matrix means small changes in $\mathbf{b}$, caused by noise, yield large changes in the solution $\mathbf{x} = H^{-1}\mathbf{b}$. In this section, we show how constrained solutions can handle these problems.

In most images, the data vary smoothly except at isolated points or edges. For example, an image of a star field will vary smoothly, except at locations of individual stars. We can make use of this image properly, smoothness, by applying a constrained least squares fit. Specifically, we minimize a linear operator $\|Qx\|$ (i.e. the sum of the squares in Qx), where Q is a matrix designed to control smoothness or other characteristics of the image (Twomey 1963, Philips 1962). For example, we can control smoothness in the one dimensional case by minimizing the second difference in the solution subject to some other constraint. If the statistical properties of the noise are known, we could minimize the second difference such that the norm of $\|Hx - b\| = $ n; that is to say, the difference of the blurred image and the solution reconvolved with the point spread function should have the same properties as the noise. In this case (minimize the second difference), Q would have the form:

$$Q = \begin{pmatrix} 0 & 0 & & & & \\ -1 & 2 & -1 & & & \\ & -1 & 2 & 1 & 0 & \\ & & -1 & 2 & -1 & \\ & & & 0 & 0 & \end{pmatrix} .$$

We use the method of Lagrangian multipliers, sometimes called the method of undetermined multipliers, to compute a solution, $\mathbf{x}$, given by (Andrews 1977):

$$\mathbf{x} = (H^T H + \gamma Q^T Q)^{-1} H^T \mathbf{b} \tag{5}$$

γ is the reciprocal Lagrangian multiplier which can be selected to control the smoothness of the solution. Solutions using Lagrangian multipliers place no restrictions on the form of Q. This flexibility allows the development of a variety of constraints depending on the known properties of the image.

Figure 1 shows the application of this constrained least squares filter for a test case (a point source) with different values of γ_2. The subscript 2 is used to indicate that the constraint is the minimum second difference. Note in figure 1.c. with the largest value of γ_2, noise in the solution has been supressed. However, the width of the point-source profile is almost as wide as the blurred profile. Also, some ringing in the restored profile is evident. Restored values on each side of the profile drop significantly below the background level. These problems result

because the second difference is large at the location of a point source. We therefore remove the second-difference constraint at the point source by setting the rows of Q corresponding to the point-source location to zero. Figure 1.f shows a restoration of the same test image when the second difference constraint is not applied at the point source. A significant improvement is apparent.

A direct extension of the method to two dimensional images is to minimize the Laplacian at each point. The Laplacian operator has a value at each pixel equal to four times the pixel value minus the values of the four immediate neighboring pixels. We use the subscript, L, to indicate the presence of the Laplacian constraint. As before, we set rows of the matrix Q to zero when the Laplacian constraint is not appropriate (i.e. edges or point sources).

The constraint need not be binary: we can vary the amount of constraint between no constraint to full constraint for any pixel, simply by multiplying the appropriate row in Q by a constant factor running from 0 to 1.

Another useful constraint is to minimize the difference of $\mathbf{x}$ from a trial solution (i.e. minimize $\|\mathbf{p} - \mathbf{x}\|$). The solution using Lagrangian multipliers is given by (Twomey 1963):

$$\mathbf{x} = (H^T H + \gamma_t I)^{-1}(H^T \mathbf{b} + \gamma_t \mathbf{p}) \quad (6)$$

where $\mathbf{p}$ is the trial solution, I is the identity matrix, and γ_t is the reciprocal Lagrangian multiplier. The subscript, t, will be used to identify the constraint as minimization of the solution from a trial solution. Some possible choices for the trial solution, $\mathbf{p}$, are a constant value (i.e. all zeros) or the blurred image itself. In either case, the ill-conditioned nature of H can be avoided and reasonable solutions obtained.
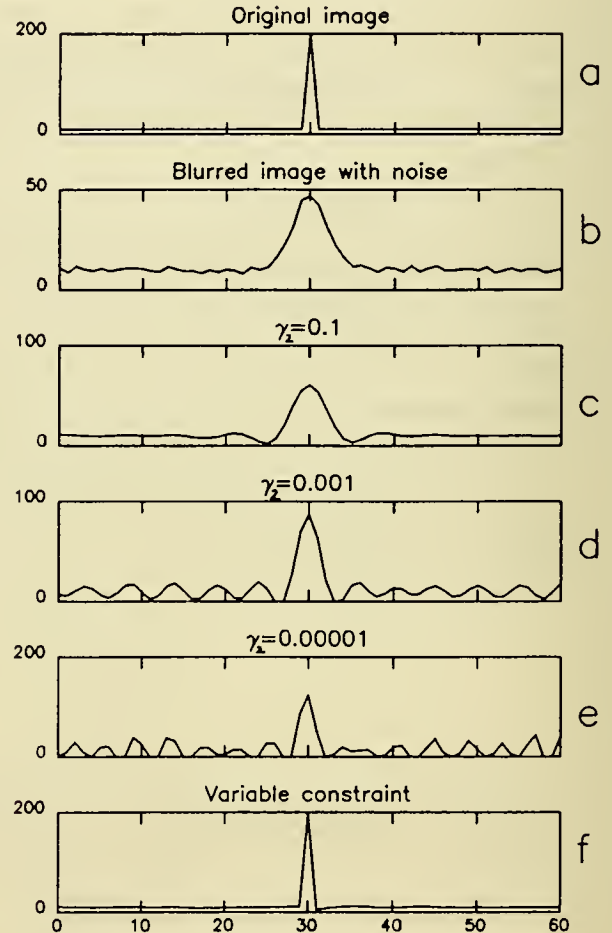

Figure 1. Effect of Langrangian multipliers. (a) original image;
(b) image blurred with a Gaussian PSF (σ=2.0 pixels) and noise added (σ=1 DN);
(c) restoration with γ_2=0.1;
(d) restoration with γ_2=0.001;
(e) restoration with γ_2=0.00001;
(f) restoration with γ_2=0.1 with constraint removed at the point source.

Multiple image constraints can be applied simultaneously:

$$\mathbf{x} = (H^T H + \gamma_a Q^T Q + \cdots + \gamma_t I)^{-1}(H^T \mathbf{b} + \gamma_t \mathbf{p}) \qquad (7)$$

where a different value of γ can be selected for each constraint.

Selection of the reciprocal Lagragian multipliers can be done by visual inspection of the results for various values or by examination of the difference of blurred image and the solution re-convolved with the point spread function. This difference should have the same properties as the noise.

2.3 Missing or Bad Data Values

A problem occurs when trying to restore images with missing or bad data values (*i.e.* cosmic ray hits or bad CCD columns). If these defects are not taken into account in the restoration, their bad values will propagate to a larger portion of the output solution. (To some extent, every point in the solution depends on all values in the blurred image.)

One method of handling bad pixels is to attempt to repair them before restoration by interpolating from neighboring values. This approach is successful only if the repair is accurate. An alternative method is to make no attempt at prior repair but handle them in the restoration process. In this approach, the restored image will have more data values than the blurred image, and the linear system is underdetermined and, therefore, singular (*i.e.* no direct inverse exists). To ignore defective pixels, we set the corresponding rows in matrix H to zero.

This method of implementation (as opposed to removing row H creating a non-square underdetermined system) allows us to keep the matrix H square and decrease the complexity of implementation. Keeping H square in no way alleviates the problem of singularity. However, the method of constrained least squares solution does alleviate the problem of singularity and obtains reasonable solutions.

3. RESULTS

3.1 Ground Based image of QSO 2130+099

Although the first example is of a ground base image, it illustrates a case that may come up frequently with Hubble Space Telescope (HST) data. We have a bright point source on a lower level diffuse source. The wings of the bright source makes it difficult to study the underlying diffuse structure. One question, which may prove crucial to the understanding of the origin of QSO's, is: What kind of galaxy plays host to a QSO? The difficulty in answering this question is that the host galaxy appears as a faint fuzz around the bright QSO.

Our approach to deconvolve the QSO image is to make the assumption that the center of the galaxy contains a point source. As described in section 2.2, we use the method of constrained least-squares, apply two constraints simultaneously, one involving smoothness in the restored image (eq. 5), the other involving the deviation from a trial solution (eq. 6). The smoothness constraint is appropriate for the host galaxy and background sky. It is empatically not appropriate for the nucleus (QSO), since that by definition is a point-source. This is where
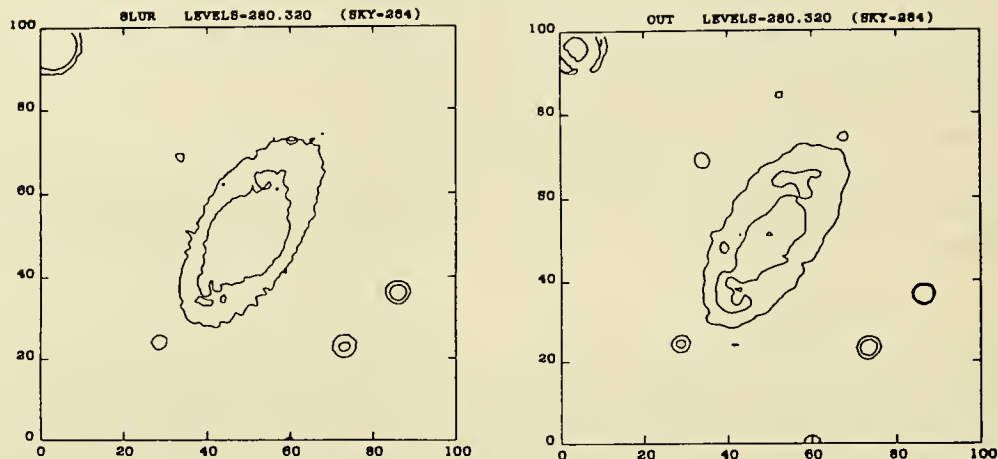
BLUR LEVELS=280,320 (SKY=254)   **OUT LEVELS=280,320 (SKY=254)**

Figure 2.   Contour plots of Quasar 2130+099. *Left*: the original image. The maximum count (at the nucleus of Q2130+099) is 21900 counts/pixel. *Right*: the restored image.   The count level at the nucleus (the quasar) is now 324,000 counts/pixel. Both images have a plate scale of 0.6 arcsec per pixel; thus both span a 1' x 1' field.   The two contour levels are at 280 and 320 counts per pixel; the average sky level is 254 counts/pixel.
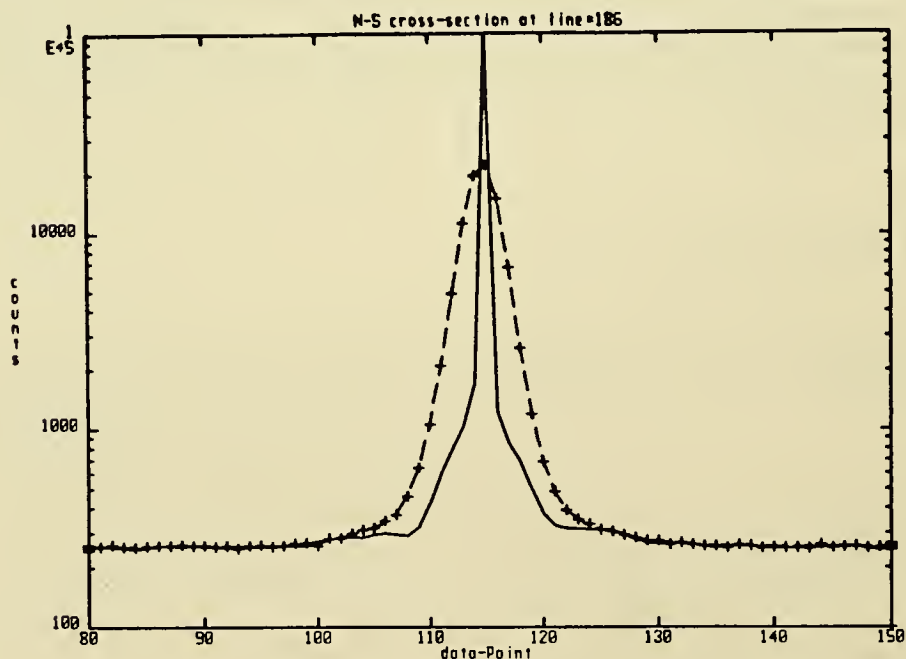

Figure 3.   Cross-sectional plot of Q2130+099.   The solid line shows the restored image. (The maximum count actually goes off-scale to 324,000 counts.) The pluses show the original (blurred image), while the dashed curve shows the result of convolving the restored image with the point-spread function.

the algebraic approach is so useful: it allows us to have local control of the constraints. We apply the smoothness constraint (minimize the Laplacian) and trial solution constraint (trial solution = sky background) to all the pixels in the image except the QSO.

Our example is a CCD image of QSO 2130+099 obtained by Tim Heckman at the 4-meter telescope at CTIO. QSO 2130+099, also known as II Zwicky 136, is a relatively nearby QSO with a redshift of only 0.06. Figure 2-left shows a contour plot ot the observed image at its lowest count levels. Not only can we see the galaxy extending 0.5 arcmin across the sky, but we can see two protrusions from the nucleus that look like spiral "stumps" if not full spiral arms. What we seek from deconvolution of Q2130+099 is not so much to enhance the resolution as it is to remove the veiling of the host galaxy by the QSO. Ideally, we would like to suck up all the flux from the QSO (nucleus) into a single pixel, so that we can look at what is around it. Figure 2-right shows the contour plot of the restored image at the same contour levels as before. Now the spiral arms arms are more prominent and fully develloped.

Figure 3 shows a cross-sectional plot of the restored image. It shows Q2130+099 for what it is: an exceedingly bright nucleus (324,000 counts) embedded in a galxy whose surface brightness falls off exponentially with increasing distance from the center, a brightness distribution typical of spirals.

3.2 Wide Field Camera Image of R136

Figure 4-left shows a Wide Field Camera image of R136 in the Large Magellanic Cloud. It shows a crowded field of stars embedded in an underlying halo resulting from the wings of the HST point spread function. We could use the approach of the previous example: measuring the locations of all of the stars and applying a smoothness constraint at all locations except the star locations. This, however, would be difficult for very close stars and very dim stars which are difficult to see in the image. Errors in the stars' locations will result in a solution with artifacts. Another approach is to decrease the constraint at pixels with a higher probability of containing a star. A simple measure of the probability is the flux in the pixel. The larger the flux, the more probable that the pixel contains a star. In the solution. (figure 4-right) we have used a weighted constraint that minimizes the norm (sum of the squares) of the difference of the solution from zero. The weight at each pixel was selected as the log of its value in the previous iteration divided by its value in the previous iteration. As the solution converges the constraint converges to the minimization of:

$$\sum_i x_i log(x_i) \tag{8}$$

Note that the solution is no longer linear. Brighter stars are much sharper than dimmer stars. This can cause problems when performing photometry in the restored image. The accuracy of the photometry in the restored image is of major importance and will require additional investigation before we can make a reasonable assessment of the success of our restoration.
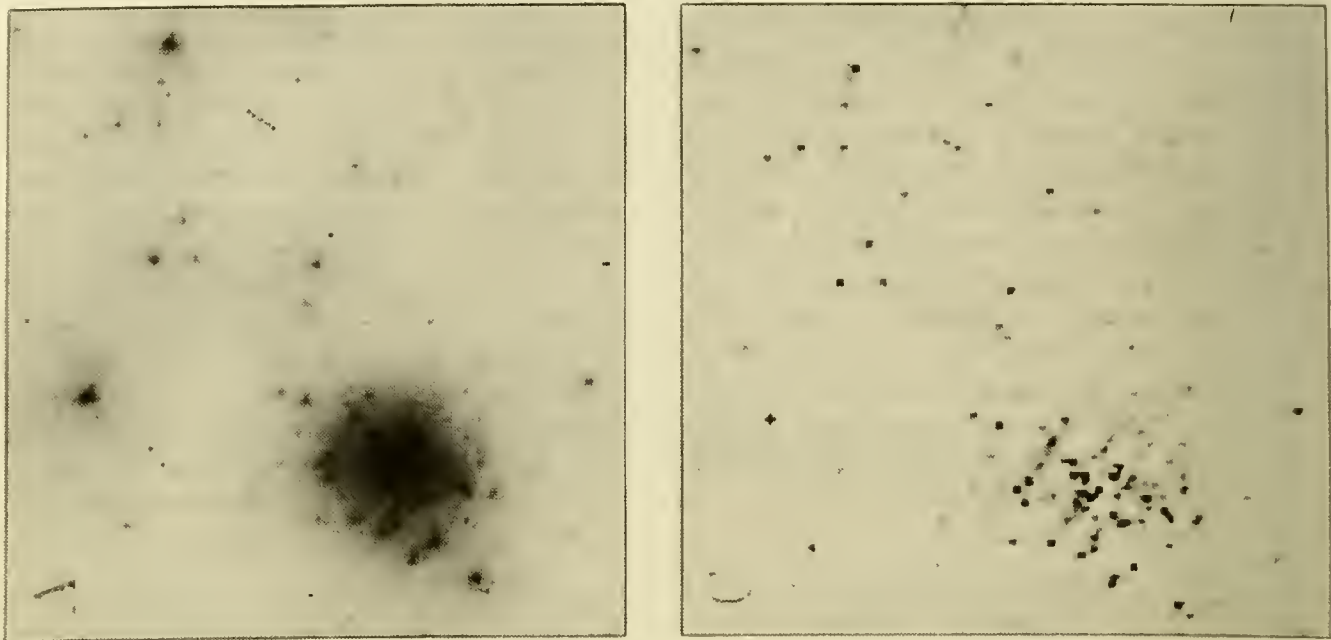
Figure 4: Restoration of Wide Field Camera image of R136

3.3 Wide Field Camera Images of Saturn

Figure 5 shows the restoration results for Wide Field Camera images of Saturn taken with three different wavelength filters. The images on the left are the unrestored, bias-subtracted and flat-fielded images. The images on the right show the results of the block iterative restoration algorithm using the minimization of the Laplacian ($\gamma_L=0.001$) and the minimization of the difference of each iteration from the previous iteration ($\gamma_t=0.01$) to constrain the solution. No attempt was made to repair bad pixels (e.g. cosmic rays) in the raw data. Instead, their locations were manually flagged and treated as missing data as described in section 2.3.

3.4 Goddard High Resolution Spectograph

The last example is a one-dimensional spectrum taken by the Goddard High Resolution Spectrograph (GHRS). The GHRS has two square science apertures: a 2.0 x 2.0 arcsecond Large Science Aperture (LSA) and a 0.25 x 0.25 arcsecond Small Science Aperture (SSA). The apertures have the feature that they cut off the wings of the HST point spread function at the edges of the apertures. The result is that the HST spherical aberration causes almost no loss of resolution when the SSA is used but does cause a significant loss of light. Only approximately 15 percent of the light from a point source centered in the SSA passes through the aperture.
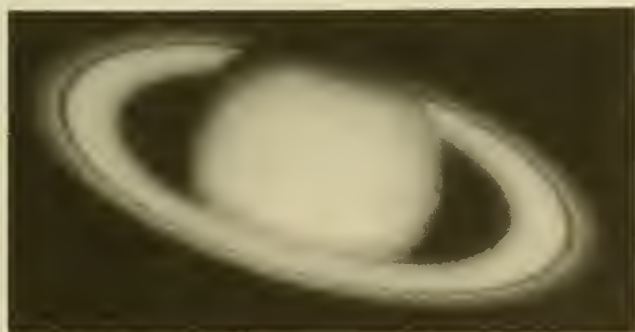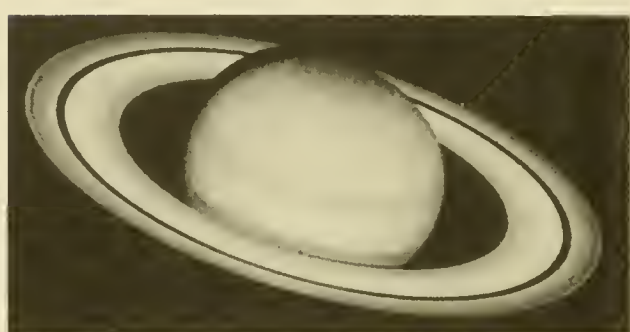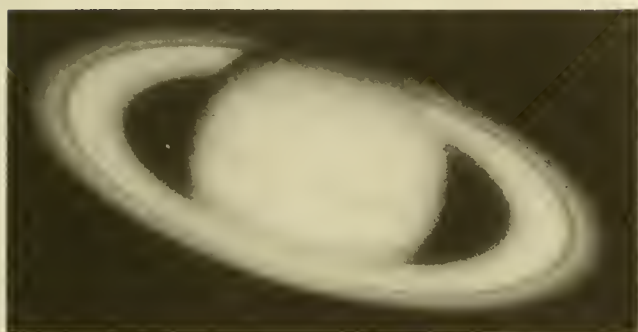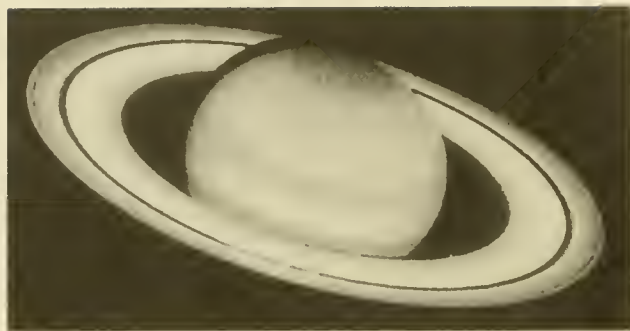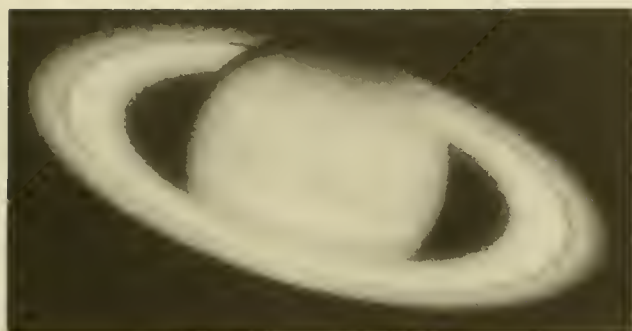
Figure 5: Restorations (right) of bias subtracted and flat-fielded WFC images of Saturn. Top - filter F439W. Middle - filter F547M, Bottom - filter F718M
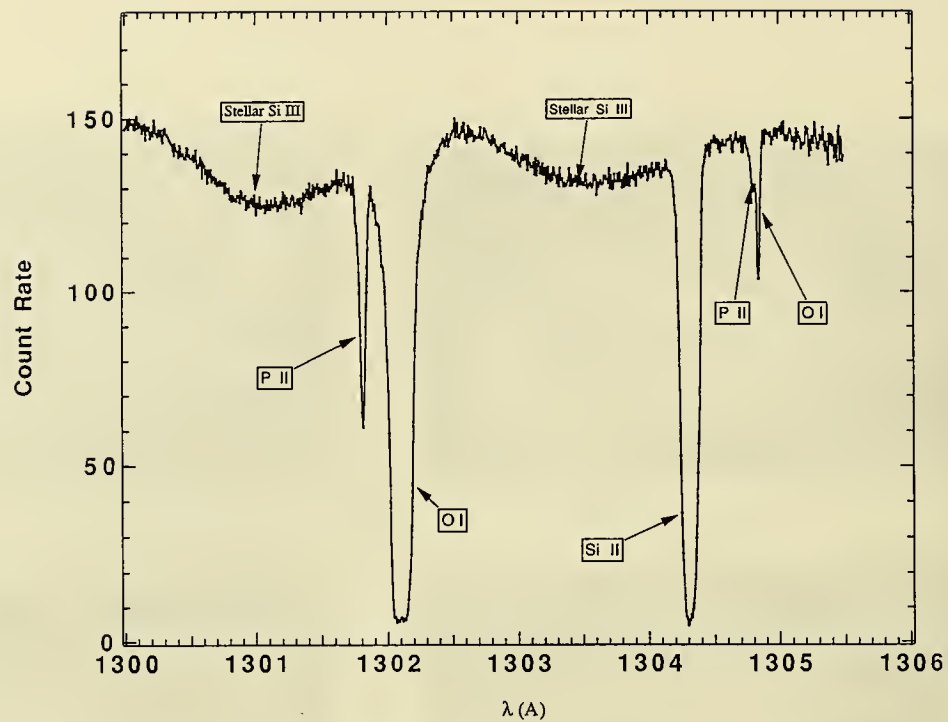
Figure 6: GHRS Large Science Aperture spectrum of ξ Persei
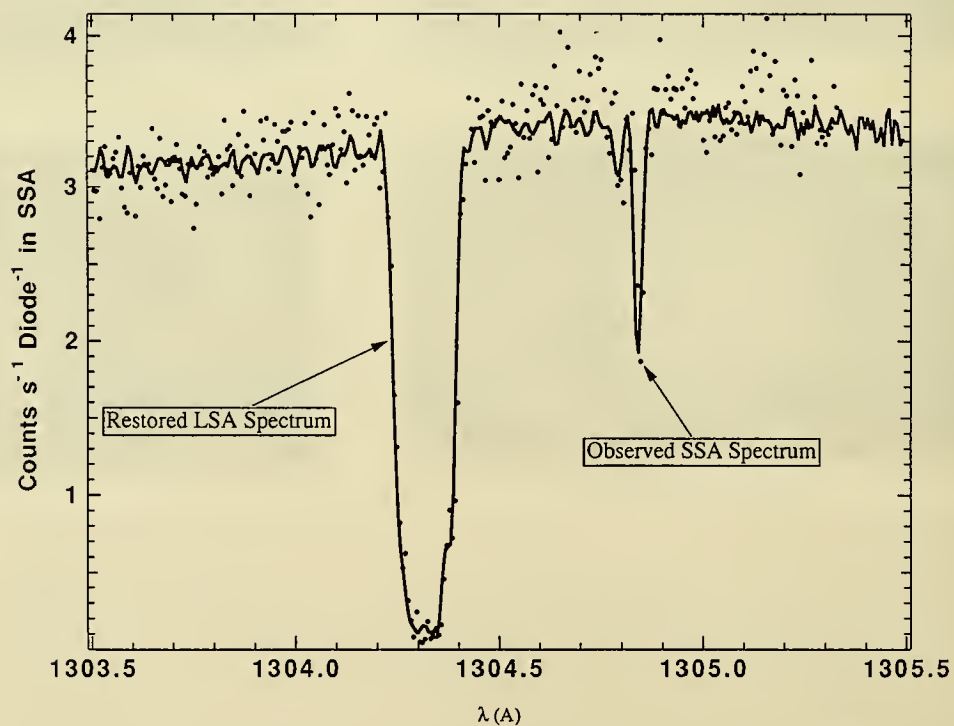
Figure 7: Comparison of the restored LSA spectrum (solid line) with an observed SSA spectrum (dots)
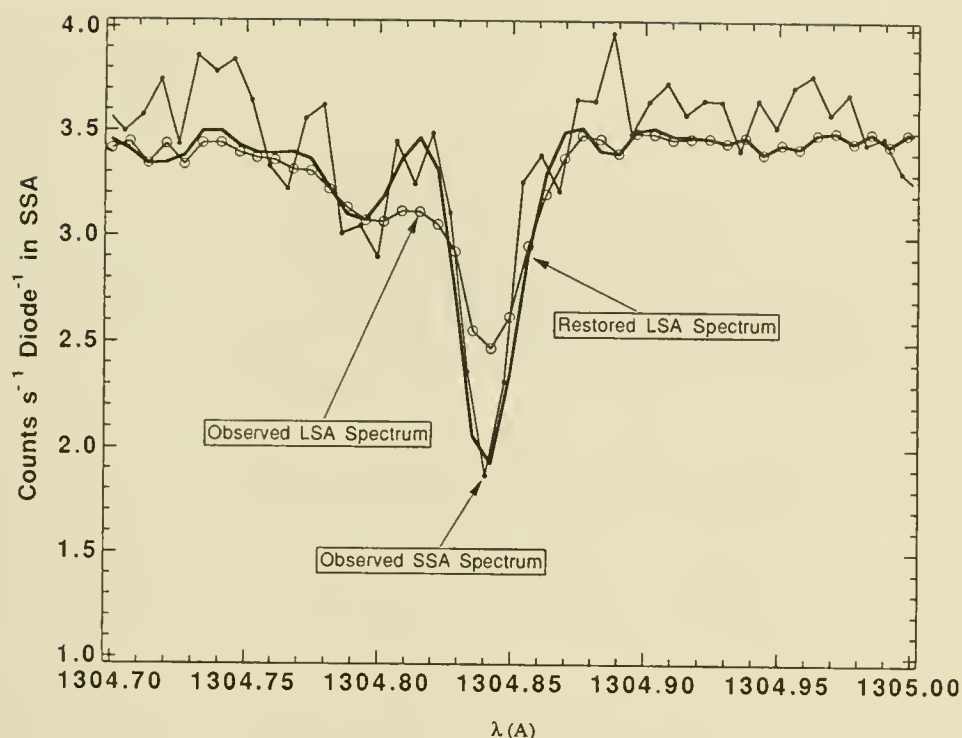
Figure 8: Comparison of GHRS restored and unrestored LSA spectrum with a SSA spectrum

Approximately 60 percent of the light passes through the LSA at the cost of a significant amount of resolution.

A user of the GHRS must use the SSA (with a significant light loss) to obtain the best resolving power unless deconvolution of LSA spectra can recover the resolution of SSA spectra. Figure 6 shows a spectrum of ξ Persei observed through the LSA. Figure 7 shows the results of the block iterative restoration (solid line) compared to a SSA spectrum with a much smaller signal to noise (dots). Not only does the restoration separate the P II and O I lines at approximately 1304.8 angstroms (Also see figure 8) but it correctly restores the profile of the Si II line at 1304.3 angstroms.

REFERENCES

Andrews. H. C., Hunt, B. R. 1977, *Digital Image Restoration* (Prentice Hall: New Jersey). pp. 148-149.

Philips. D. L. 1962. "A Technique for the Numerical Solution of Certain Integral Equations of the First Kind", *J. ACM*. **9**. 84-97.

Twomey, S. 1963, "On the Numberical Solution of the Fredholm Integral Equations of the First Kind", *J. ACM*,**10**. 97-101 (1963).

Young, D. M. 1971, *Iterative Solution of Large Linear Systems*, (Academic Press: New York). pp. 434-437.

Deconvolution of Simulated HST Frames by MIM

J.Pfleiderer, Institute of Astronomy, A-6020 Innsbruck, Austria

Abstract: The minimum-information method (MIM) is based on a diophantic approximate solution of a linear system of equations. The data is convolved with the PSF and then deconvolved with the PSF convolved with itself. The quadratic smoothness constraint - minimization of structural information - is equivalent to an unconstrained deconvolution by a Prussian-helmet type modification of the deconvolving function. MIM compares favorably with other methods, as MEM, CLEAN, inverse or Wiener filter, or Lucy's algorithm. Advantages are a good resolution for point sources and extended sources, good recognition of weak extended features, and a linear response that makes photometry rather safe. Possible applications include images, visibilities, spectra, and time series. I give examples with a simulated HST image.

Introduction

While some astronomers still consider deconvolution as sufficiently unreliable to be not worthwhile at all or even deceiving, others have argued that they would not believe in any deconvolved feature not recognizable in the original frame. I shall show that the first argument is wrong (but nevertheless good for damping exaggerated optimism) while the second one is, within certain limits, reasonable and helpful. Deconvolution can, or rather should, only make more apparent those pieces of information that already exist in the data.

The method

The theory of MIM (minimum information method) was given by Pfleiderer (1985). It is based on the notion that a deconvolution should draw as little information (in the general information–theoretical sense of low probability or "seldomness") as possible not only from the (unknown) noise but also from the (as well unknown) "true" (i.e. noiseless) data. This information is, of course, smallest whenever the true data correspond to what is expected from the deconvolved image. Accordingly, MIM uses a least-squares fit of discrete noiseless data by a discrete set of point sources (assuming a known PSF). Such fit results in a linear system of equations that contain an (unknown) noise term. That is, the system is not to be solved rigorously but rather only approximately within the noise level. The point sources are determined in integer multiples of the detection limit which effectively avoids noise-fitting. As smoothness constraint MIM uses minimization of "structural information" as defined by Pfleiderer (1990). Variable smoothing is possible and even asked for by the theory. This feature is particularly useful for deconvolution of bright stars. A final additional smoothing within the noise level removes the coarseness of the result.

The set of equations is

$$\sum_{j} a^*_{jk} f_j - b_k = n_k \quad ; \qquad j, k = 1, .., J. \tag{1}$$

Here, f is the intensity (integer numbers before the final smoothing) of the deconvolved image at image pixel j, b is the data convolved with the normalized PSF at image pixel k. The PSF and the data must be given on or interpolatable to that grid.

The term n_k is a normalized noise term. In the simplest case, it is expected to be a Gauss distribution with average 0 and variance 1. That is, the f's should be chosen such that the left sides of the equations deviate from zero by a number that fits into the expected distribution.

$$a^*_{jk} = a_{jk} + \Gamma g^*_{jk} \tag{2}$$

is the effective deconvolving beam where a is the normalized PSF convolved with itself, and Γ is a Lagrange parameter which may in the general case depend on j. Contrarily to other methods, the Lagrange parameter

and thus the degree of smoothing is not fixed by the theory. It can be chosen according to the problem. g_{jk}^* is derived from the term g_{jk} in the above-mentioned "structural information"

$$S = \sum_{j}\sum_{k}(f_j - f_k)^2 g_{jk} \qquad (3)$$

which is to be minimized. Its quadratic form in f assures that it enters the set of equations in linear form. This is similar to the "Smoothness-Stabilized CLEAN" (Cornwell 1983) and can be interpreted as an unconstrained deconvolution with a Prussian-Helmet type "beam" a^* rather than the original "beam" a.

The constraint function g_{jk} depends on the distance between the positions of the pixels j and k. Its value is large for distances smaller than the PSF width, and small for large distances. Thus, g represents essentially a local smoothing constraint. This is a unique feature of MIM: Deconvolution and smoothing is local. The equations do not contain any global term, a feature particularly useful for achieving good resolution. Sufficient sampling and intensity provided, resolution to better than the width of the PSF (superresolution) is generally no problem.

Most methods have global components: The entropy is a global constraint, Fourier coefficients are globally influenced by all data. That is, the deconvolution at one position depends on what is happening elsewhere. This makes quantitative estimates of the deconvolved intensities rather difficult. MIM, as a local linear method, is thus particularly adapted to doing photometry.

On the other hand, global components are useful to make the result of a deconvolution unique. The MIM result is not unique, the non-uniqueness being, however, essentially restricted to the noise level. While some theorists consider uniqueness a mandatory feature of any good deconvolution method, I do not think so. Any solution that fits the data and obeys certain constraints is as good as any other - none is definitely superior. MIM can handle positive and negative intensities, the latter being sometimes useful for radioastronomical absorption features. However, a non-negativity constraint is often useful and can be easily implemented into the solving procedure of eq.(1).

Another convenient feature of MIM is that it can handle asymmetric PSFs. A good example is seismic data where one analyses damped oscillations. Here the "PSF" is totally one-sided and has essentially zero sum.

Isolated point sources

When a star is deconvolved by a smoothing method, it is represented by an extended intensity distribution that will not reproduce the data. The residuals (observed data minus expected data) will be positive in the center and have a negative ring around it. This can be partly remedied by a ring of negative intensity (or, more strictly, by rings of alternatingly positive and negative intensities). Every smoothing method must produce either bad residuals or these Gibbs-like rings. A non-negativity constraint is of little help against rings if the star stands atop of extended emission.

Stimulated by the discussions on HST data, I have recently incorporated into MIM a search for stars which are then deconvolved without smoothing. Fig.3 shows that this new feature does, in its present version, still miss or partly miss some stars which are accordingly deconvolved with rings but, on the other hand, interprets some narrow structures as a combination of point sources and extended sources.

The need for such a procedure in MIM can be seen as follows: A good data fit asks for $\sum af \approx b$, the introduction of the smoothing constraint for $\sum a^* f \approx b$, i.e., $\sum \Gamma g f \approx 0$. Thus, large intensities f ask for a small Γ, or little smoothing.

HST and deconvolution

The PSF of the HST seems to consist of a small central peak approximately as wide as originally planned but containing only one fifth or so of the intensity it should have contained. This peak is surrounded by a large halo. Such form of the PSF implies that the achievable resolution is only slightly decreased. If the HST were up to the plans, one would be able to do some superresolution. As it stands now, one probably should be content with recovering - more or less - the originally planned resolution.

The intensity contained in the halo can to a high degree be removed by deconvolution but, by that, it does not contribute to the information on the image. That is, the loss of about 2 magnitudes in sensitivity for point sources cannot be recovered. Also, some loss in dynamic range is unavoidable. Deconvolution can improve the dynamic range but must necessarily fail, due to the unavoidable photon noise, to recover it to the same degree as it recovers the high-intensity features, be they stars or extended emissions.

A special problem for the WF/PC is the varying PSF. The theory of MIM does not pose problems in handling varying PSFs but the computational effort would much increase.

Varying PSFs can often be approximated by the convolution of an extended "blurred point image" with an average (i.e., constant) extended PSF. It would be worthwhile to try what a well-chosen constant PSF will do to the WF/PC frames.

Examples

ESA/ST/ECF kindly provided simulated HST data by overlaying a smoothed image of the galaxy M101 at its $\approx$ 150fold distance over a star map (Plejades at their $\approx$ 800fold distance) to produce a field of about 11x11 arcsec (512x512 pixels of size 22x22 mas) for the FOC. It was convolved with a PSF resembling an HST PSF but being rotationally symmetric (fig.1), to give fig.2. This was MIM-deconvolved (fig.3) assuming constant noise giving an SNR per pixel of about 200 in the center of the galaxy. Even faint stars of the Plejades and the faint structures of M101 are well reproduced. Some bright stars are deconvolved into one pixel and are, thus, not easily recognized in the figure. The straight lines are easily avoidable artifacts (the deconvolving function was taken too small, only about ± 2.2 arcsec = ± 100 pixels). These were, however, only detected in the final representation.

Fig.2 had also been scaled to exposure times of 1000, 10000, and 30000 sec of which I show only the first one, having about 300 photons per pixel in the bright center of the galaxy. Photon noise was correspondingly added (fig.4 with SNR$\approx$17 per pixel in the center but $\leq$ 5 in most regions). The MIM deconvolution is
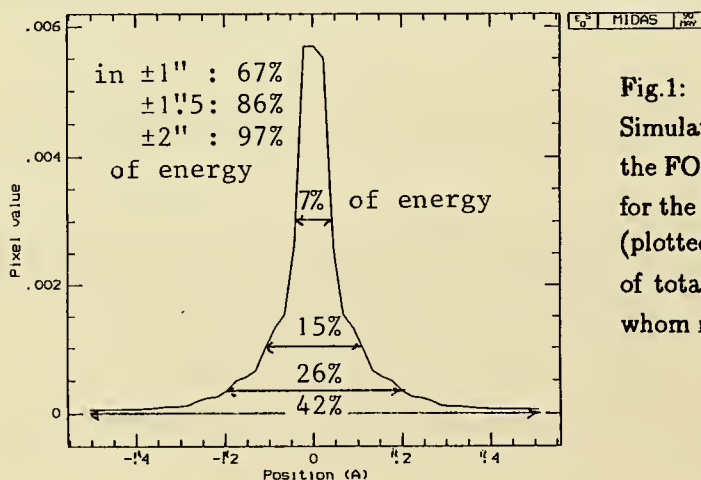
Fig.1:
Simulated HST PSF (point spread function) for the FOC (f:96) but rotationally symmetric except for the innermost pixels. Scales: Abscissa: arcsec (plotted range ±0".5). Ordinate: part (per pixel) of total energy. Provided by ESA/ST/ECF to whom my thanks are due.

MIDAS version: 90MAY

Fig.3: MIM deconvolution of fig.2. Scale: pixels. Assumed
pixel noise is about 0.5% of the central galaxy brightness.
For the rings around stars and the (avoidable) straight-
line artifacts, compare text. Note that the white areas
are within the assumed noise level of the lowest gray step,
so they are conspicuous but of low significance.

MIDAS version: 90MAY

Fig.2:
Simulated HST frame: M101 + Plejades scaled down to
11.2x11.2 arcsec = 512x512 pixels of 22x22 mas each
and convolved with PSF of fig.1. Scale: arcsec. Noise-
less except for rounding errors. Provided by ESA/ST/ECF.

Fig.5:
MIM deconvolution of fig.4. Weak stars are not recovered
from the noise, and the extended features are less sharp
than in fig.3.

Fig.4:
Fig.2 with photon noise added. About 300 photons per pixel
in bright center of galaxy, corresponding to a SNR per
pixel of about 17. The SNR per pixel is less than 5 for
80% of all pixels. Provided by ESA/ST/ECF.

54

Fig.6:
MIM. Enlarged map section of fig.5 (in a slightly
different version), corresponding to fig.1 of Adorf
et al. (this volume). The brighter stars are rec-
ognized as such and partly deconvolved as point
sources (one or two black pixels), the fainter ones
are deconvolved as extended (and therefore weak)
features, the faintest ones are lost, depending on
the local noise.

shown in fig.5. Even here, the resolution for stars is better than 0.1 arcsec. Many weak stars are lost in the
noise but most extended features are recovered. This is because the effective SNR for extended objects is
much larger than the SNR of single pixels.

Not all deconvolved features can be seen in the data reproductions of figs. 2 and 4 but all are traceable in
the data. Nevertheless, some improvements in the MIM deconvolution routine are possible and would be
advantageous.

Conclusions

HST data can be effectively improved by deconvolution. Any statement to the contrary is plainly wrong. I
consider this as one of the undisputable results of this workshop.

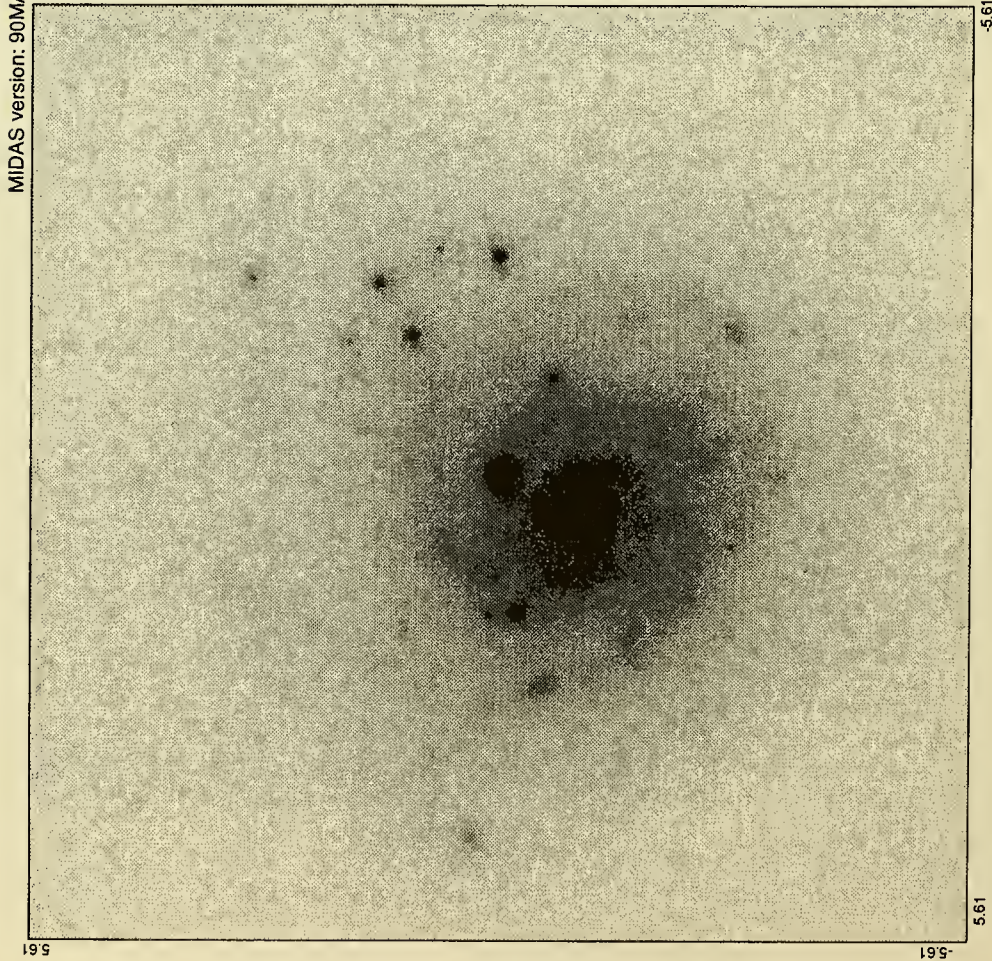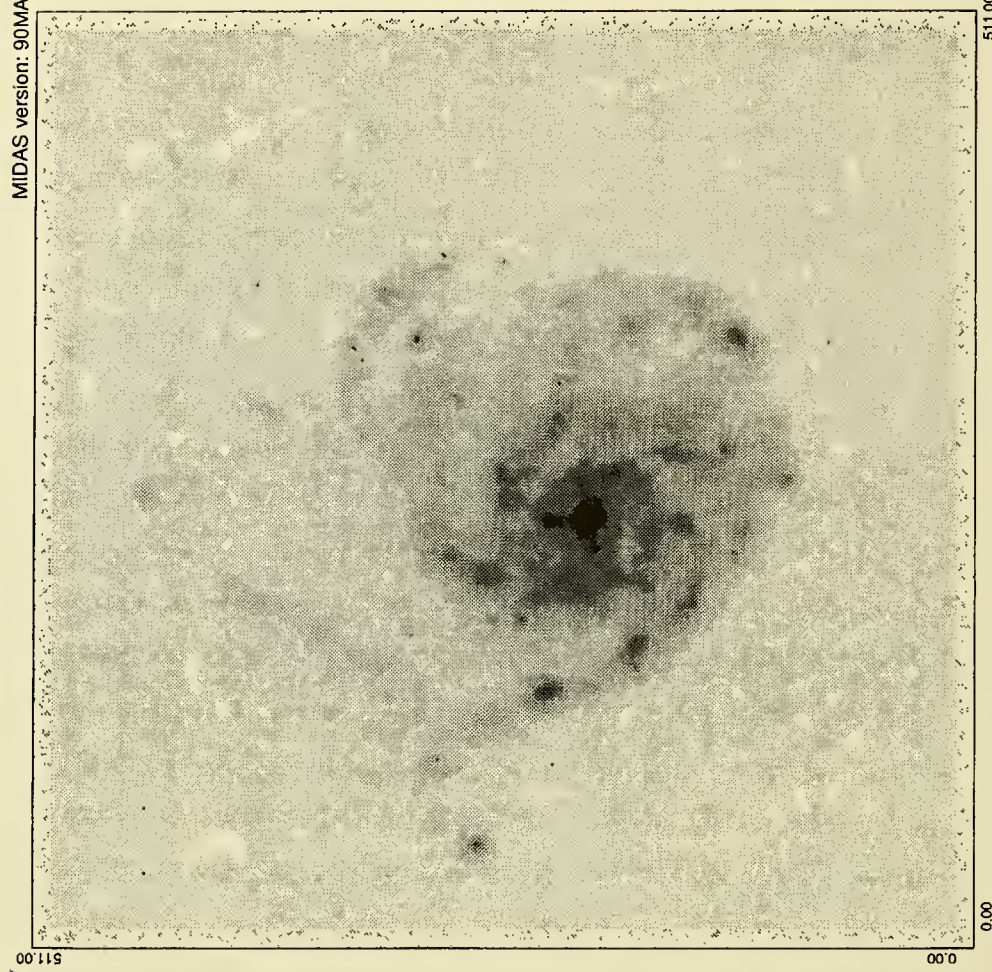Undersampling (as it occurs in the WF/PC) can be overcome to some degree if the PSF is known on a finer
grid, e.g., by a combination of several slightly shifted undersampled PSFs. MIM is a strong deconvolution
method that in all tests I could as yet do performed at least as good as any other method. This is illustrated
by fig.6 as compared to fig.1 of Adorf et al. (1991). A good test example (as that given here) deconvolved by
different methods allows a detailed comparison and shows advantages and disadvantages of these methods
or their present computer representations, thus contributing to the improvement of deconvolution routines.
A difficulty lies in assuring that the results are truly comparable in the premises and in the representations.
I think, however, that one should not just stick to one method but rather use several ones in order to extract
a maximum of information from the HST data.

References:

Adorf H.-M., Walsh J.R., Hook R.N. 1991: This volume

Cornwell T.J. 1983: *Astron. Astrophys.* **121**, 281

Pfleiderer J. 1985: *South Afr.J.Phys.* **8**, 73

Pfleiderer J. 1990: p.197 in: *Errors, Bias and Uncertainties in Astronomy*, (C.Jaschek,F.Murtagh eds.),
        Cambridge Univ. Press; *Naturw.* **76**, 297, 1989

Modifications of the Lucy-Richardson Iteration for Restoring Hubble Space-Telescope Imagery[1]

Donald L. Snyder
Electronic Systems and Signals Research Laboratory
Washington University
St. Louis, Missouri 63130

1. Abstract

The relationship of the Lucy-Richardson iteration method for restoring images to methods used for forming images of radionuclide distributions is identified. On the basis of this relationship, several modifications of the Lucy-Richardson iteration are suggested. These include flat-field corrections for detector nonuniformity, corrections for background radiation and read-out noise, and the inclusion of regularization to suppress noise and edge artifacts.

2. Introduction

In the early 1970s, L. Lucy [2] and W. Richardson [7] published an iterative method for recovering images of objects from optical data in which the objects are blurred. Their method has been rediscovered more recently by L. Shepp and Y. Vardi [11] in the context of medical imaging where the spatial concentration of a radioactive tracer within the body is sought from noisy data in which the concentration is blurred. The connection between these two seemingly disparate fields of science, restoring optical imagery and radionuclide imaging, occurs because they are stochastic inverse-problems having a common mathematical description involving quantum-limited data. The purpose of this paper is to note some implications for restoring optical imagery of developments in radionuclide imaging where more recent and powerful mathematical tools have been applied successfully [4,6,12,14]. These developments yield extensions of the Lucy-Richardson iteration method that may be important for restoring images acquired with the Hubble space-telescope.

3. The Stochastic Inverse-Problem

The Lucy-Richardson iteration solves the following stochastic inverse-problem. With reference to Fig. 1, there is a field of objects that emit incoherent light with a space-dependent intensity that will be denoted by $\lambda(x)$. This is a nonnegative function having the interpretation that its integral over any region is the average optical-energy emanating from that region and received at the entrance pupil of the telescope. Light is gathered with the telescope and focused onto a detector array, resulting in the production of photoelectrons. The photoconversion intensity will be denoted by $\mu(y)$, which is determined ideally by the object intensity and the point-spread function $p(y \mid x)$ of the telescope according to

$$\mu(y) = \beta \int_x p(y \mid x)\lambda(x)\,dx, \tag{1}$$

where $\beta = \eta/h\bar{\nu}$ is the ratio of the quantum efficiency (η) of the detector and the average energy ($h\bar{\nu}$) in an incident photon. The point-spread function is nonnegative, space dependent, and to a good approximation given by the squared magnitude of the Fourier transform of the aperture function, accounting for obscurations present in both the OTA and relay optics and weighted by a phase function that accounts for spherical aberration, with the superimposed obscurations of the OTA and relay optics varying with position in the detector array. I assume without loss of generality that the PSF, $p(y \mid x)$, is normalized as a probability density, $\int_y p(y \mid x)\,dy = 1$. Modifications of this intensity to account for flat-field variations, background radiation, and read-out noise are discussed below.

Photoconversions in the detector array are modeled as a spatial Poisson-process $N(\cdot)$ with intensity function $\mu(y)$. Thus, the number of photoelectrons in a pixel of the array is a Poisson-distributed random variable with a mean equal to the integral of $\mu(y)$ over the pixel. The locations of photoelectrons and their number form the data that are acquired; these data are quantized by the detector array into numbers of counts in each pixel of the array.

The deblurring problem is to estimate the intensity distribution $\lambda(x)$ given the quantum-limited data $N(\cdot)$ and subject to the constraint that $\lambda(x) \geq 0$. The log-likelihood functional of the Poisson-distributed data is given by [15]

56

Object Space: X   PSF: p(y|x)   Data Space: Y
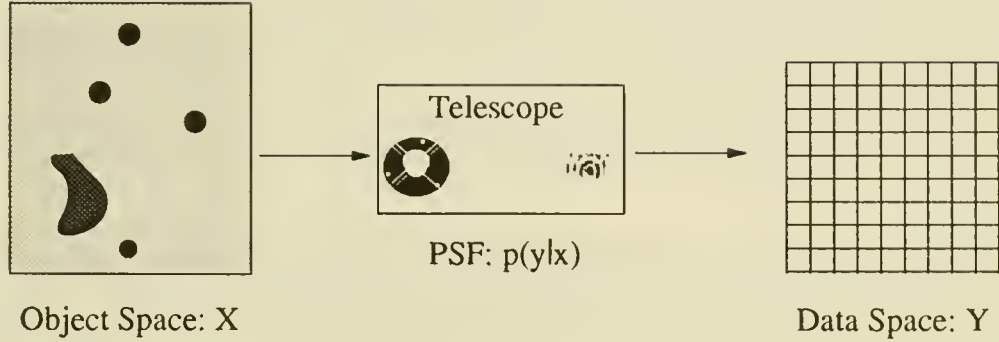
Figure 1

$$\mathcal{L}(\lambda) = -\int_X \lambda(x)\,dx + \int_{\mathcal{Y}} \ln\left[\int_X p(y \mid x)\lambda(x)\,dx\right] N(dy). \qquad (2)$$

This quantity is maximized, subject to the nonnegativity constraint, for the maximum-likelihood method of solving this problem, which is a well established statistical approach.

3.1 The Shepp-Vardi Iteration

In 1982, L. Shepp and Y. Vardi [11] derived a numerical solution to the above deblurring problem in the context of radionuclide imaging. In their approach, the estimate of $\lambda(x)$ is selected to maximize the likelihood (2) of the Poisson-distributed data. Their numerical solution is derived using the expectation-maximization algorithm of A. Dempster, N. Laird, and D. Rubin [1] and is of the following iterative form: .

$$\hat{\lambda}_{k+1}(x) = \hat{\lambda}_k(x) \int_{\mathcal{Y}} \left[\frac{p(y \mid x)}{\int_X p(y \mid x')\hat{\lambda}_k(x')\,dx'} \right] N(dy). \qquad (3)$$

When the object and data spaces are discrete, or pixelized, this integral expression becomes

$$\hat{\lambda}_{k+1}(x) = \hat{\lambda}_k(x) \sum_{\mathcal{Y}} \left[\frac{p(y \mid x)}{\sum_X p(y \mid x')\hat{\lambda}_k(x')} \right] N(y), \qquad (4)$$

in which the indices x and y represent pixel locations in the object X and data $\mathcal{Y}$ spaces, respectively, $N(y)$ is the number of photoelectrons in detector pixel y, and $\lambda(x)$ is the mean number of counts in pixel x. The discretizations of the two spaces do not need to be the same, so subpixelization schemes where X is partitioned more finely than $\mathcal{Y}$ are included, but having more unknown values in X space than known values in $\mathcal{Y}$ space will require regularization, which is needed in any case as discussed below. The convergence of the iteration (4) towards a maximum-likelihood estimate of $\lambda(x)$ has been established [16,17].

The Shepp-Vardi solution (4), based on maximizing likelihood for Poisson-distributed data, is identical to the Lucy-Richardson [2,7] iteration. This iteration has the properties that the estimates at each stage of the iteration satisfy the nonnegativity constraint, $\hat{\lambda}_k(x) \geq 0$, and the total energy is preserved, $\int_x \hat{\lambda}_k(x)\,dx = N(\mathcal{Y})$. The operations in (4) are the following, starting with the denominator on the right side: (1), blur the stage-k estimate with the PSF; (2), divide the result into the data pixel by pixel; (3), blur the result with the adjoint PSF; (4), multiply by the stage-k estimate pixel by pixel to get the stage-$(k+1)$ estimate; and (5), repeat until convergence. These operations are readily implemented for a space invariant PSF; then, step-1 requires a convolution with the PSF, and step-3 requires a convolution with the PSF rotated by 180 degrees.

The modifications of the Lucy-Richardson iteration that are summarized below were derived in the references cited by maximizing likelihood using the expectation-maximization algorithm, following the Shepp-Vardi procedure. I write the modifications in the integral form (3), which includes the discrete form (4) when the spaces are pixelized.

4. Modifications

4.1 Correction for Flat Field, Background Radiation, and Read-Out Noise

Elements in the detector array exhibit nonuniformity of response. Let $\beta(y)$ denote the response of the detector at y; this is the fraction of incident optical photons converted into photoelectrons. Moreover, additional photoelectrons due to background radiation and thermoelectrons due to dark current may be present. In this case, the intensity function (1) of the Poisson process modeling the data becomes

$$\mu(y) = \beta(y)\int_x p(y \mid x)\lambda(x)\,dx + \mu_0(y), \tag{5}$$

where $\mu_0(y)$ is the background plus thermoelectron intensity. The problem of forming the maximum-likelihood estimate of $\lambda(x)$ given Poisson-distributed detector data with intensity (5) and the constraint $\lambda(x) \geq 0$ is addressed by D. Politte and D. Snyder [6]. The iteration (3) becomes

$$\hat{\lambda}_{k+1}(x) = \hat{\lambda}_k(x)\frac{1}{\overline{\beta}(x)}\int_{\mathcal{Y}}\left[\frac{\beta(y)p(y \mid x)}{\int_x \beta(y)p(y \mid x')\hat{\lambda}_k(x')\,dx' + \mu_0(y)}\right]N(dy), \tag{6}$$

where $\overline{\beta}(x) = \int_{\mathcal{Y}}\beta(y)p(y \mid x)\,dx$ is the fraction of photons emanating from x that are converted to photoelectrons anywhere in the detector array. It is evident that the nonnegativity constraint is satisfied at each iteration stage, but energy normalization is not maintained. In the limit as $k \to \infty$, there holds

$$\int_x \hat{\lambda}_\infty(x)\overline{\beta}(x)\,dx = N(\mathcal{Y})\left\{1 - \int_{\mathcal{Y}}\left[\frac{\mu_0(y)}{\beta(y)\int_x p(y \mid x')\hat{\lambda}_\infty(x')\,dx' + \mu_0(y)}\right]N(dy)\right\}. \tag{7}$$

There is an issue needing to be addressed about read-out noise present with the CCD detector array. The modeling of this noise as Gaussian distributed and additive complicates the data model and invalidates the Poisson model leading to the Lucy-Richardson iteration. However, the small level of this noise (≈ 13 electrons/pixel r.m.s.) indicates that it can be accommodated as Poisson distributed and included in the dark-current parameter $\mu_0(y)$. In this way, (5) includes the effects of both background radiation and read-out noise.

4.2 Regularization

The problem of estimating $\lambda(x)$ as stated above is ill posed with the result that solutions produced with the Lucy-Richardson iteration (4) exhibit instabilities that become more and more pronounced as iterations proceed. Regularization is needed to overcome this undesirable effect. In [14], we have introduced the method of sieves and resolution kernels for regularization. With this method, an estimate is sought not for $\lambda(x)$ but, rather, a blurred version $d(w) = \int_x r(w \mid x)\lambda(x)\,dx$ that we term the *desired function*. Here, $r(\cdot)$ is a resolution kernel that must be selected by the user; this is a nonnegative function normalized as a probability density. The idea is that estimating $\lambda(x)$ to arbitrarily fine resolution is not supported by the finite amount of data available, so we seek to estimate this intensity up to some finite resolution that is. A natural choice for $r(\cdot)$ would be the diffraction-limited PSF that would have

occurred in the absence of spherical aberration, but performance studies might indicate, as they have for radionuclide imaging, that a somewhat greater sacrifice in resolution than this "intrinsic resolution-limit" may in fact be necessary. Further regularization is introduced to control the destabilizing effects of quantum noise, which appear as speckle in the estimate. This is accomplished with a sieve to constrain the estimate to be a member of a smooth set of functions $S = \{d(\cdot); \ d(w) = \int_z s(w \mid z) \xi(z) dz\}$, where $\xi(\cdot)$ is an arbitrary intensity and $s(\cdot)$ is a given sieve-kernel, which is a nonnegative function normalized as a probability density [13,14]. The two kernels, $r(\cdot)$ and $s(\cdot)$, can be selected arbitrarily - of course, the performance of the estimator is dependent on the choice - subject to the requirement that they must be such that a nonnegative function $q(\cdot)$ exists as a solution to the integral equation

$$\int_w q(y \mid w) r(w \mid x) dx = \int_w p(y \mid w) s(w \mid x) dw. \tag{8}$$

The function $q(\cdot)$ becomes the kernel in the Lucy-Richardson iteration, the final iterate of which is blurred by the sieve kernel to produce the regularized estimate of $d(w)$. The determination of $q(\cdot)$ is in general difficult, but there are two cases where this is straightforward. The first is when the PSF, resolution kernel, and sieve kernel are all Gaussian functions, as discussed in [14]. The second is when the sieve and resolution kernels are selected to be the same; then, $q(\cdot)$ equals the PSF $p(\cdot)$.

The estimation problem becomes: given data governed by the Poisson model with parameter (1), and given the resolution and sieve kernels, estimate the desired function $d(\cdot)$ subject to the constraint that the estimate be a member of the sieve S. The result is the following modified form of the Lucy-Richardson iteration when the sieve and resolution kernels are identical:

$$\hat{\lambda}_{k+1}(x) = \hat{\lambda}_k(x) \int_{\mathscr{Y}} \left[\frac{p(y \mid x)}{\int_x p(y \mid x') \hat{\lambda}_k(x') dx'} \right] N(dy), \quad \hat{d}_k(w) = \int_x s(w \mid x) \hat{\lambda}_k(x) dx. \tag{9}$$

Thus, for the choice of equal kernels, one performs the usual Lucy-Richardson iteration and then blurs the final iterate by the kernel of the sieve. After extensive studies with real and simulated data, this is the choice we use for radionuclide imaging. This result has been extended to include flat-field correction, background radiation, and read-out noise [6].

Another form of regularization can also be used, which I mention because of the connection that results between the Lucy-Richardson iteration and maximum-entropy methods. This is introduced by adding a penalty function to the log-likelihood (2) so it becomes

$$\mathcal{L}(\lambda) = - \int_x \lambda(x) dx + \int_{\mathscr{Y}} \ln \left[\int_x p(y \mid x) \lambda(x) dx \right] N(dy) + \alpha S(\lambda), \tag{10}$$

where α is a Lagrange multiplier, and $S(\lambda)$ is a given functional of the intensity. The first two terms on the right are a result of the Poisson statistics describing photoconversions and replace the integral of the squared difference of the data and intensity (1) that occurs with white Gaussian-noise statistics. The last term may also be interpreted as assuming that the intensity has prior statistics described by a Gibbs' distribution $Z^{-1} \exp[\alpha S(\lambda)]$. The maximum-entropy method for Poisson-distributed data results from the choice $S(\lambda) = \int_x \lambda(x) \ln \lambda(x) dx$. One now estimates $\lambda(x)$ by maximizing (10) subject to the nonnegativity constraint. The result of doing so with the expectation-maximization algorithm results in the following modified form of the Lucy-Richardson iteration:

$$\hat{\lambda}_{k+1}(x)[1 - \alpha \delta S(\hat{\lambda}_{k+1})] = \hat{\lambda}_k(x) \int_{\mathscr{Y}} \left[\frac{p(y \mid x)}{\int_x p(y \mid x') \hat{\lambda}_k(x') dx'} \right] N(dy), \tag{11}$$

where $\delta S(\lambda)$ is the functional variation of $S(\lambda)$. For example, the left-hand side for the maximum-entropy method becomes $-\hat{\lambda}_{k+1}(x) \ln \hat{\lambda}_{k+1}(x)$, and (11) must then be solved at each iteration for the stage-$(k+1)$ estimate, for example by using a Newton iteration nested within each stage of the Lucy-Richardson iteration. References [13] and [10] can be consulted for more details on this approach to regularization, which is now being studied intensely for radionuclide imaging.

5. Additional Comments and Discussion

5.1 Performance Evaluation for Photometry

Analytical prediction of the quantitative accuracy of estimates produced via the Lucy-Richardson iteration is difficult and largely remains an intractable problem. For this reason, we have performed extensive experiments in the use of this method for radionuclide imaging [4,5,6,14]. These experiments have been both by computer simulations and by collecting and processing real data from known distributions of radioactivity. Results were evaluated quantitatively for bias and variance, and they were evaluated qualitatively by experienced radiologists and neurologists who regularly view reconstructions of radioactivity distributions. As a result of these studies, we are currently using a modified Lucy-Richardson iteration that includes regularization with equal resolution and sieve kernels and compensation for detector nonuniformity and background, as in (6) and (9) combined, for imaging positron-emitting radionuclides because we have found that it produces more quantitative reconstructions than alternative methods, including one based on solving the normal equations via regularized pseudo-inverses.

A similar performance study will be needed, with practical parameter values and levels of photon-flux, to state with confidence that the Lucy-Richardson iteration can produce quantitative restorations of Hubble space-telescope imagery so that accurate photometry can be performed. However, due to the similarity of the mathematical models for the two stochastic inverse-problems, our experience suggests strongly that the modified Lucy-Richardson iteration will outperform methods based on solving linear matrix-equations (i.e., the normal equations) via regularized pseudo-inverses.

5.2 Computational Issues

The Lucy-Richardson iteration is a very demanding computation. We normally perform 200 iterations of the modified algorithm in order to reach a stable solution for radionuclide imaging; we have preferred to continue iterations until convergence rather than terminate them at some intermediate stage where the optimality properties of the estimate are undefined. This requires so much computation time that the method is regarded as impractical for radionuclide imaging when a conventional single-stream processor is used. For this reason, we have studied special processing architectures to perform the task practically. Two multiprocessor approaches are presently being used. One is based on transputer technology [8,9], and the other is based on the Discrete Array Processor (DAP) made by Active Memory Technology, Inc., which is an SIMD mesh-connected array of single-bit processors [3,10]. Both architectures make it possible to exploit the substantial potential for parallelism offered by the modified Lucy-Richardson iteration so that reconstructions are produced in practical times.

Four graduate students in our laboratory (C. Butler, A. Hammoud, S. Joshi, and A. McCarthy) and I tried to implement the regularized Lucy-Richardson iteration (9) on an AMT DAP/610 multiprocessor, which has 4096 processing elements, using two byte integer arithmetic and a 512×512 image array. Simulated data were created using (5) as the intensity for Poisson-distributed counts including read-out noise; the space varying PSF was computed using the Fourier method mentioned above. We failed to produce restorations of the simulated data in the limited time we had available to implement the simulation and try the method, but our experience may be enlightening. We did not have enough memory (10 to 20 gigabytes might be needed) to store the space variant PSF and adjoint PSF needed to perform the iteration (6), and computing the PSF within each stage of the iteration proved to take too long to get anywhere near the 200 iterations that may be needed. While our implementation was on a powerful parallel machine and drew upon our experience in similar computations for radionuclide imaging, it is very likely that a more efficient implementation can be identified to reduce the computation time, but our initial effort indicates that the restoration problem for the Hubble space-telescope is very demanding indeed if the space-variant PSF is to be taken into account fully. On the other hand, the restoration problem is quite manageable if a space-invariant PSF is assumed. Thus, a design that is a compromise between a fully space-variant and fully space-invariant PSF needs to be accomplished to realize a practical solution having acceptable restoration and photometry.

6. Conclusions

Data acquired with the wide-field, planetary, and faint-object cameras of the Hubble space-telescope appear to be well modeled by a Poisson counting process having an intensity that is the intensity distribution of the astronomical field blurred by the point-spread function of the telescope optics. Application of the expectation-maximization algorithm for maximizing likelihood numerically results for this model in the Lucy-Richardson iteration. It follows, therefore, that to the extent the Poisson model is accurate, the Lucy-Richardson iteration, and its modifications given herein, are mathematically fundamental and natural for restoring Hubble space-telescope imagery.

7. Acknowledgement

8. References

1. A. D. Dempster, N. M. Laird, and D. B. Rubin, "Maximum Likelihood from Incomplete Data via the EM Algorithm," *Journal of the Royal Statistical Society*, B, Vol. 39, pp. 1-38, 1977.

2. L. Lucy, "An Iterative Technique for the Rectification of Observed Distributions," *The Astronomical Journal*, Vol. 79, pp. 745-754, June 1974.

3. A. McCarthy, r. Barrett, and M. I. Miller, "Systolic Implemenation of the EM Algorithm for Emission Tomography on Mesh Connected Architectures," Proc. 1988 Conf. on Information Sciences and Systems, Department of Electrical Engineering, Princeton Univ., Princeton, NJ, pp. 702-707, March 1988.

4. D. G. Politte and D. L. Snyder, "The Use of Constraints to Eliminate Artifacts in Maximum-Likelihood Image Estimation for Emission Tomography," *IEEE Transactions on Nuclear Science*, Vol. 35, pp. 608-610, February 1988.

5. D. G. Politte, "Image Improvements in Positron-Emission Tomography Due to Measuring Differential Time-of-Flight and Using Maximum-Likelihood Estimation, *IEEE Transactions on Nuclear Science*, April 1990.

6. D. G. Politte and D. L. Snyder, "Corrections for Accidental Coincidences and Attenuation in Maximum-Likelihood Image Reconstruction for Positron-Emission Tomography," *IEEE Transactions on Medical Imaging*, to appear 1990, and Research Monograph ESSRL-89-12, Electronic Systems and Signals Research Laboratory, Washington Univ.

7. W. H. Richardson, "Bayesian-Based Iterative Method of Image Restoration," *Journal of the Optical Society of America*, Vol. 62, pp. 55-59, January 1972.

8. F. U. Rosenberger, G. C. Johns, D. G. Politte, and C. E. Molnar, "Transputer Implementation of the EM Algorithm for PET Image Reconstruction," Proc. Third Conf. of the North American Transputer Users Group, Sunnyvale, CA, pp. 51-62, April 1990. (Available also as C.S.L. Tech. Note 337, Inst. for Biomedical Computing, Washington Univ., St. Louis, MO, Feb. 1990)

9. F. U. Rosenberger, G. C. Johns, D. G. Politte, and C. E. Molnar, "Optimizing Communications for a Transputer Implementation of the EM Algorithm for PET Image Reconstruction," IEEE Nuclear Science Symp., Arlington, VA, October 1990.

10. B. Roysam, J. A. Shrauner, and M. I. Miller, "Bayesian Imaging Using Good's Roughness Measure - Implementation on a Massively Parallel Processor," Proc. IEEE International Conf. on Acoustics, Speech, and Signal Processing, New York, pp. 932-935, March 1988.

11. L. A. Shepp and Y. Vardi, "Maximum-Likelihood Reconstruction for Emission Tomography," *IEEE Transactions on Medical Imaging*, Vol. MI-1, pp. 113-121, 1982.

12. D. L. Snyder and D. G. Politte, "Image Reconstruction from List-Mode Data in an Emission Tomography System Having Time-of-Flight Measurements," *IEEE Trans. on Nuclear Science*, Vol. NS-30, pp. 1843-1849, 1983.

13. D. L. Snyder and M. I. Miller, "The Use of Sieves to stabilize Images Produced with the EM Algorithm for Emission Tomography," *IEEE Transactions on Nuclear Science*, Vol. NS-32, pp. 3864-3872, October 1985.

14. D. L. Snyder, M. I. Miller, L. J. Thomas, Jr., and D. G. Politte, "Noise and Edge Artifacts in Maximum-Likelihood Reconstructions for Emission Tomography," *IEEE Transactions on Medical Imaging*, Vol. MI-6, pp. 228-238, September 1987.

15. D. L. Snyder and M. I. Miller, *Random Point Processes in Time and Space*, Springer-Verlag, in preparation for publication in 1990. (This is a revision of D. Snyder, *Random Point Processes*, Wiley, 1975.)

16. Y. Vardi, L. A. Shepp, and L. Kaufman, "A Statistical Model for Positron Emission Tomography," *Journal of the American Statistical Association*, Vol. 80, pp. 8-35, March 1985.

17. C. F. J. Wu, "On the Convergence Properties of the EM Algorithm," *The Annals of Statistics*, Vol. 11, pp. 95-103, 1983.

ITERATIVE MAXIMUM LIKELIHOOD ESTIMATOR AND BAYESIAN ALGORITHMS FOR IMAGE RECONSTRUCTION IN ASTRONOMY

Jorge Llacer
Engineering Division
Lawrence Berkeley Laboratory
University of California
Berkeley, CA 94720
and
Jorge Nuñez
Departament d' Astronomia
Universitat de Barcelona and Observatorio Fabra
Barcelona, Spain

Abstract

This paper will discuss Maximum Likelihood Estimator and Bayesian methods of image reconstruction as applied to Astronomy. The motivation for the algorithms is given in terms of the statistical distribution of the data to be reconstructed. Target functions and iterative formulas for Poisson distributed data and a new iterative method for Poisson data with additive Gaussian noise will be described. The concept of feasible images will be discussed in some detail and the expected benefits of using reconstruction methods yielding such images will be illustrated by describing some results obtained in medical tomography.

1. INTRODUCTION

The problem of image reconstruction in Astronomy is fundamentally identical to that of reconstructing images obtained from Emission Tomography (ET) in Nuclear Medicine. The data come from different sources: high energy photons (γ-rays) emitted by disintegrating radioisotopes introduced as tracers into the human body generate the data in ET, while relatively low energy photons from celestial sources generate the signals detected in Astronomy. In both cases there is a blurring function, which in the case of ET can modify the data sufficiently to render the emitting structures unrecognizable in the unreconstructed data. Most of the work discussed in this paper has had its origin in ET and it is possible that the terminology used may not always coincide with the language of Astronomy. A proper definition of terms should avoid confusion. This paper will first define the problem of image reconstruction in general terms, describe what a Maximum Likelihood Estimator (MLE) is and how it is used in image reconstruction, indicate the desirable characteristics of MLE images as well as their shortcomings, describe the concept of "feasibility" and introduce the concept of Bayesian reconstructions. Within the Bayesian case, the Fast Maximum a Posteriori method with Entropy prior (FMAPE) algorithm will be described and its use leading to feasible images will be shown. Finally, a new algorithm for the reconstruction of Poisson distributed data in the presence of additive Gaussian noise will be proposed. Image examples from medical tomography will not be included in this paper, but a substantial number of references will be collected for those readers interested in details of that area of work.

2. STATEMENT OF THE PROBLEM

We introduce first the notation that will be used in this paper:

p_j $j = 1,...,D$ - the projection data (measured data)

a_i $i = 1,...,B$ - activity in the source, or emission density (parameters to be estimated)

f_{ji} - the transition or point spread matrix (probability that an emission in pixel i in the source will be detected in pixel j of the data

$h_j = \sum_{i=1}^{B} f_{ji} a_i$ - the forward projection or blurring operation

$\mathbf{p, a, f, h}$ - the corresponding vectors/matrix

Then, the reconstruction problem consists in estimating the values a_i from a set of measurements p_j by minimizing some measure of the residuals

$$p_j - h_j$$

in accordance with some rules, applying some constraints. In order to minimize the residuals, a "target function" is maximized or minimized. This function defines the rules and may include the constraints. It is tempting to write the imaging equation, i.e., the equation that defines the problem to be solved, in a typical manner as

$$\mathbf{fa + n = p} \tag{1}$$

where $\mathbf{n}$ is a noise vector. We will refrain from doing that because (1) cannot cover the case of Poisson distributed data $\mathbf{p}$. Indeed, if that is the case, the statistical distribution of the noise vector $\mathbf{n}$ is undefined.

3. THE MAXIMUM LIKELIHOOD ESTIMATOR

A. Gaussian Statistics

In general terms, a Maximum Likelihood Estimator (MLE) is a procedure for estimating some parameters from a set of measurements by maximizing a function that defines the likelihood of obtaining those measurements from the parameters. In particular, for Gaussian distributed variables with a single variance σ^2, the Likelihood function is

$$L(\mathbf{a}) = P(\mathbf{p}|\mathbf{a}) \propto \prod_{j=1}^{D} e^{-\frac{(p_j - h_j)^2}{2\sigma^2}} \tag{2}$$

where

$$h_j = \sum_{i=1}^{B} f_{ji} a_i . \tag{3}$$

$P(\mathbf{p}|\mathbf{a})$ is the probability of obtaining a vector of measurements $\mathbf{p}$ given a source vector of mean activities $\mathbf{a}$. Maximization of (2) is equivalent to minimizing $\log L$, i.e.,

$$\log L = \sum_{j=1}^{D} \frac{(p_j - h_j)^2}{2\sigma^2} \tag{4}$$

Minimization of (4) is equivalent to finding a minimum least squares solution to the imaging problem for a fixed, unspecified σ. Thus, a MLE solution for Gaussian variables with fixed σ is equivalent to a minimum least squares solution. It can be shown that image reconstruction by "inverse filtering" or "deconvolution" using Fourier transforms results in a minimum least squares solution of minimum norm,[1] with possibly other constraints related to limited bandwidth. It follows that the "deconvolution" method of image reconstruction is equivalent to a MLE solution for Gaussian variables of fixed σ, with some norm and frequency spectrum contraints.

The emission of photons from celestial sources obeys Poisson statistics, which only approaches a Gaussian distribution in the limit of large number of photons. The detection process is fundamentally a "thinning" process that preserves the Poisson statistics, although it generally reduces the mean. Under conditions of low number of detected photons, it would appear that a Likelihood function based on Poisson statistics could lead to better reconstructions than one based on Gaussian statistics. This has been found to be the case in ET and the same can be expected in Astronomy. A measure of what a "better" reconstruction means will be discussed below.

B. Poisson statistics

A method for computing a MLE solution to the ET problem was published in 1982 by Shepp and Vardi.[2] Their method is based on the Expectation Maximization (EM) algorithm of Dempster, Laird and Rubin.[3] The target function that is being maximized is

$$L(\mathbf{a}) = P(\mathbf{p}|\mathbf{a}) = \prod_{j=1}^{D} e^{-h_j} \frac{h_j^{p_j}}{p_j!} \tag{5}$$

where h_j is defined in (3). The iterative formula resulting from the maximization can be written as

$$a_i^{(k+1)} = a_i^{(k)} \sum_{j=1}^{D} \frac{f_{ji} p_j}{\sum_{l=1}^{B} f_{jl} a_l^{(k)}} \tag{6}$$

where the superscript (k) indicates the value at the kth iteration. We have shown that the only initial guess for **a** which is consistent with no prior knowledge (in addition to the detected data) about the image to be reconstructed is a uniform image field.[4] The use of the unreconstructed image in Astronomy as an initial guess will be discussed below in the section on feasible images.

The iterative formula (6) is identical to Lucy's method of image restoration[5] which was derived from Bayesian considerations, without specifically referring to Poisson distributions. Lucy was able to show that when applying a small fraction of the correction steps implied by (6), the likelihood increased at each iteration. The later work of Shepp and Vardi shows that the iterative process converges to a maximum likelihood estimate when applying the full correction of (6) to Poisson data.

The MLE method of (6) has a number of desirable characteristics, in addition to using the correct statistical distribution:

1) the transition matrix **f** can incorporate space-variant point spread functions. In principle, each pixel in the source could have a different point spread function.
2) Flat field corrections can be incorporated into the matrix values, so that the Poisson nature of the detected data is not disturbed by multiplication by correction factors.
3) Stellar or electronic Poisson distributed background can be subtracted by estimation. One can include one or more ficticious pixels in the matrix **f** that generate background with a probability measured experimentally or determined theoretically. The values of those pixels are estimated along with the image pixels.
4) Positivity of the results is assured.
5) With proper regularization and accurately defined transition matrices, reconstruction results have been shown to be unbiased up to the limits of the experimental method.[6,7]

The process of reconstructing images by the MLE method suffers from one particular problem whose solution has generated a substantial amount of research. As the iterative process progresses beyond a certain number of iterations, the reconstruction of distributed photon sources becomes very noisy, with high peaks and low valleys. This comes about from the unconstrained nature of the MLE reconstructions and the ill-posed nature of the image reconstruction problem, in general. Two principal solutions have been proposed: 1) regularization of the MLE by Bayesian[8,9,10] or sieve[11] methods and 2) halting the iterative process before deterioration occurs.[12,13,14] We will discuss here the latter method. Regularization by the method of sieves was presented by D.L. Snyder in the same conference where this paper was read.

4. FEASIBLE IMAGES

The likelihood function (5) will reach a maximum when all p_j are as close as possible to h_j. When used as a target function in the reconstruction problem, (6) will yield an image **a** whose projection **h** will be as close as possible to the measurements **p**, without constraints in the values of **a**. The behavior of MLE solutions at convergence is a direct consequence of this lack of constraints. Indeed, if we let the values of the elements of **h** become arbitrarily close to the corresponding elements of **p**, we are allowing the solution to follow the statistical noise too closely.

The residuals $p_j - h_j$ should not be allowed to get too small in a reconstruction, or remain too large. The final values of the residuals should be such that the projections **h** of a reconstructed image **a** could be means of Poisson distributions that have **p** as specific realization values. In other words, in the average, each value p_j should be approximately one standard deviation from the corresponding h_j. In this way, the reconstructed image would be consistent with the projection data by consideration of the statistical process of photon emission and detection.

We have called those reconstructions that fulfill the above requirements "feasible"[13] and devised a rule for determining when the iterative MLE process passes through a region of iterations that results in feasible images.[12,13]

A. Definitions

In a formal way, we use the following:

Definition 1: The image $a_1, a_2, ... , a_B$ is said to be feasible with respect to data $p_1, p_2, ... , p_D$, if and only if the statistical hypothesis that $p_1, p_2, ... , p_D$ are a Poisson sample with means $h_1, h_2, ... , h_D$, respectively, can be accepted (not rejected) at a given significance level.

An implementation of a test for feasibility has been described in Ref. 12. The first step of the test consists in scaling the differences $p_j - h_j$ for each data pair to a new variable $x(d)$ which is uniformly distributed between 0 and 1 if the data are Poisson with respect to their means. Next, a histogram with N bins is generated with the values $x(d)$ for all data pairs. Then we test the hypothesis that $x(d)$ is uniformly distributed between 0 and 1 by Pearson's procedure with N - 1 degrees of freedom. The histogram testing function H is defined as

$$H = \sum_{i=1}^{N} \frac{\left(x_i - \dfrac{D}{N}\right)^2}{\dfrac{D}{N}} \tag{7}$$

where x_i is the observed frequency of $x(d)$ falling in bin i, and D/N is the expected frequency if $x(d)$ is uniformly distributed. The above test has been found to be too stringent when dealing with real tomographic PET data. Under those conditions, we have had to devise a less restrictive test[13] that includes one adjustable parameter that corresponds to the uncertainty with which we know the transition matrix of the detecting system. The feasibility tests are based on the assumption that the projections h_j are statistically independent of the data p_j. This assumption is basically incorrect since the set **h** is obtained from the set **p** by the reconstruction process. It has been possible to show experimentally, however, that the sets **h** and **p** behave as if statistically independent for the purposes of feasibility testing.[15]

A weaker feasibility test can also be defined[13]:

Definition 2: The image $a_1, a_2, ... , a_B$ is said to be weakly feasible with respect to data $p_1, p_2, ... , p_D$, if and only if the second moments of $p_1, p_2, ... , p_D$ are consistent with the Poisson hypothesis, namely

$$\sum_{j=1}^{D} \frac{\left[p_j - h_j\right]^2}{h_j} \approx D \tag{8}$$

Indeed, the expected value of the numerator of each term in (8) is the variance, while that of the denominator is the mean. Therefore (8) must be satisfied if the Poisson hypothesis holds. However, the reverse statement is not true. This test is easier to implement than the one of Def. 1 and it is found to work just as well in ET cases in which the transition matrix is known exactly (usually in cases of simulated data). For real data, the test of Ref. 13 has had to be implemented. The uncertainty parameter has to be adjusted only once for a specific detecting instrument.

B. Obtaining feasible images

Feasible images have been obtained by a variety of methods:

1) Iterating with the MLE algorithm of (6) until the hypothesis testing function (7) takes values below ~ 36.0 for a few consecutive iterations, corresponding to a 99% confidence level.
2) Iterating past the feasibility region and filtering back into it with a small Gaussian kernel. The farther from feasibility is the iterative process allowed to go, the larger the kernel has to be. After a certain point, however, only near feasible images can be obtained. Medical images from Positron Emission Tomography (PET) obtained by iterating ~ 50% more iterations that the minimum for the onset of feasibility, with post-filtering with a small Gaussian kernel have yielded the best reconstructions that we have seen by this method.
3) From the FMAPE algorithm,[9,10] in which one adjustable parameter is selected so that the process converges to feasible images, without the image instability of the MLE. This algorithm will be described below.
4) From the method of sieves first proposed for ET by Snyder and co-workers.[11] The results of our evaluations show that the feasible images that we can obtain with real PET data (where the transition matrix is only known approximately) are not acceptable because they are too noisy. It is easy to obtain near-feasible images that are quite satisfactory, though.[16]

C. Characteristics of feasible images

An extensive statistical analysis of the characteristics of feasible images in PET has been carried out.[7] We summarize here the results obtained, which can be expected to carry over to reconstructions in Astronomy.

1) Feasible images can be sharper than "inverse filtering" images, without paying a significant penalty in noise.
2) The use of MLE images obtained by iterating the algorithm of (6) past the onset of feasibility by ~ 50% more iterations and filtering slightly results in images with no more bias than "inverse filtering" reconstructions. Similar good results can be obtained by using the FMAPE algorithm, by making it converge to a region somewhat noisier than feasible and post-filtering.
3) When an ensemble of images from independent data sets of the same source is analyzed pixel-by-pixel, feasible images have a standard deviation which is approximately proportional to the square root of the number of counts in the region considered, in contrast with "inverse filtering" images, that have a pixel-by-pixel standard deviation which is large and approximately independent of number of counts in the region.

The above results support our contention that feasible images can be "better" than "inverse filter" images for Poisson distributed data. Feasibility is not a sufficient condition for acceptance of an image, however.

D. Starting point for the iterative process

As indicated above, a uniform field is the only correct starting point for the MLE iterative process if no *a priori* information exists about the image, in addition to the initial blurred data.[4] This is due to the fact that feasible images are the result of iterating towards a maximum likelihood image (thought to be unique[17]) but one stops before maximum likelihood. At the stopping point there is a memory of the initial image and it is possible to obtain feasible images that are not acceptable if the starting point contained erroneous information.[18] It is for this reason that using the initial blurred data as the initial image $\mathbf{a}^{(0)}$ in addition to using them as detected data $\mathbf{p}$ appears unadvisable. Statistical fluctuations in the initial image will carry over to the final accepted reconstruction. The authors believe that this is a common error in the use of Lucy's algorithm. Tests will be carried out to determine the effects resulting from that practice. The FMAPE algorithm, to be described next, converges to images that are independent of the starting point, although, in the absence of prior information, the vector $\mathbf{a}^{(0)}$ should be set to a uniform field as well.[4]

5. BAYESIAN RECONSTRUCTIONS AND THE FMAPE ALGORITHM

A. Bayesian concepts

The maximization of the likelihood function $P(\mathbf{p}|\mathbf{a})$ of Eq. (5) results in an image $\mathbf{a}$ for which the measurement vector $\mathbf{p}$ has the highest probability of being the result of a measurement. A different target function could be devised that answers to the demand of obtaining the image $\mathbf{a}$ of highest probability *given* a measurement $\mathbf{p}$, i.e., maximize $P(\mathbf{a}|\mathbf{p})$. The latter target function, if it can be defined, is the correct one to maximize in the case of noisy data: Maximizing the likelihood will yield an image whose projection will follow the noisy data as closely as possible, while maximizing $P(\mathbf{a}|\mathbf{p})$ need not have that characteristic. In fact, using Bayes' rule,

$$P(\mathbf{a}|\mathbf{p}) = P(\mathbf{p}|\mathbf{a})P(\mathbf{a}) / P(\mathbf{p}) \qquad (9)$$

we see that the "Bayesian" target function $P(\mathbf{a}|\mathbf{p})$ not only contains the likelihood function, or conditional probability $P(\mathbf{p}|\mathbf{a})$, but it also incorporates $P(\mathbf{a})$, the probability distribution of the image $\mathbf{a}$, or image "prior". If we know that the image has a certain set of characteristics that can be put in the form of a probability distribution, maximization of (9) may place some limitations on how small the residuals can be, for example, resulting in a regularization of the results. This is precisely what the Fast Maximum a Posteriori algorithm with Entropy prior (FMAPE) does. Reference 9 goes into some detail into the question of justifying an entropy prior, giving adequate references. We shall only indicate here that an entropy prior results in an excellent regularization of the MLE algorithm. The form of the target function that we shall give here contains "cross-entropy" terms, which allow the incorporation of image pixel prior information into the formulation, if available, in addition to the regularizing effect of the entropy.

The target function to maximize is given by

$$\log L(\mathbf{a}) = -\sum_{i=1}^{B} \frac{a_i}{\Delta a} \log\left(\frac{a_i}{Q_i}\right) + \sum_{j=1}^{D}\left[-h'_j + \frac{p_j}{\Delta p_j}\log\left(h'_j\right)\right] -\mu\left(\sum_{i=1}^{B} a_i - N\right) \tag{10}$$

where

$$h'_j = \frac{\sum_{i=1}^{B} f_{ji} a_i}{\Delta p_j}. \tag{11}$$

Parameter Δa can be adjusted to control the relative weight of the entropy vs. likelihood in (10). If Δa is very small, entropy dominates and the method yields an image that is very close to the prior image defined by the values Q_i, or to a flat field if all Q_i were initially set to a constant in the absence of image prior information. On the other hand, if Δa is made large, the entropy term is insignificant and the solution becomes a MLE noisy result at convergence. By adjusting Δa it is possible to have the process converge to a feasible image. Figure 1 shows the values of the weak feasibility test of Def. 2 as a function of iteration number for the reconstruction of a simulated brain medical image, for different values of Δa. The region around 1.0 in the ordinate is the region of feasiblity.

In (10), the values for the data $\mathbf{p}$ are assumed to have been pre-multiplied by detector gain correction factors Δp_j. The division by those factors in (10) removes those corrections, returning the data to their Poisson nature. The division by the same factors in (11) applies the corrections to the transition matrix elements, so that the estimated values of $\mathbf{a}$ are properly corrected. The prior image pixel information Q_i can be the expected value of the mean activity in each pixel, if known. Parameter μ is a Lagrange multiplier to insure the conservation of counts in the image.
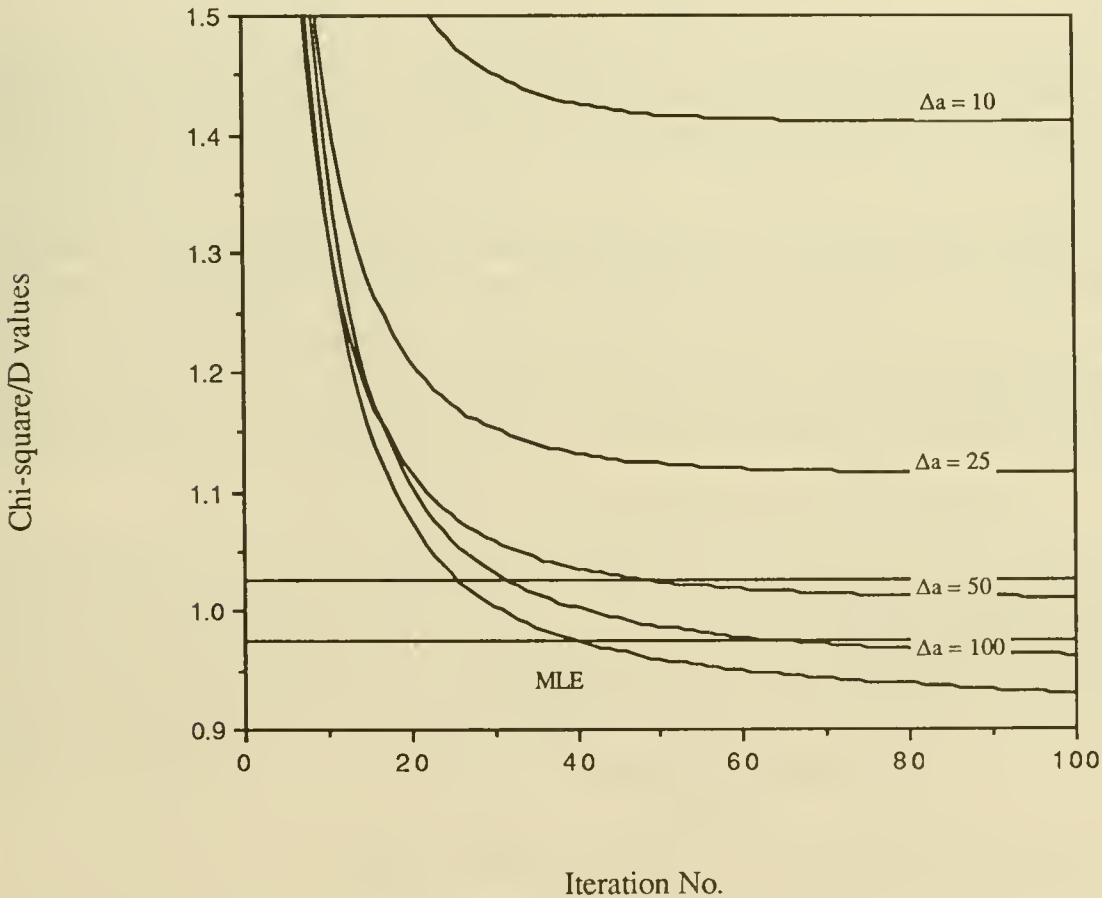
Chi-square data for Reconstructions

Fig. 1: Normalized χ^2 values for the residuals of FMAPE reconstructions of medical brain simulation data as a function of iteration number, for different values of the parameter Δa. The region in the vicinity of 1.0 corresponds to feasibility.
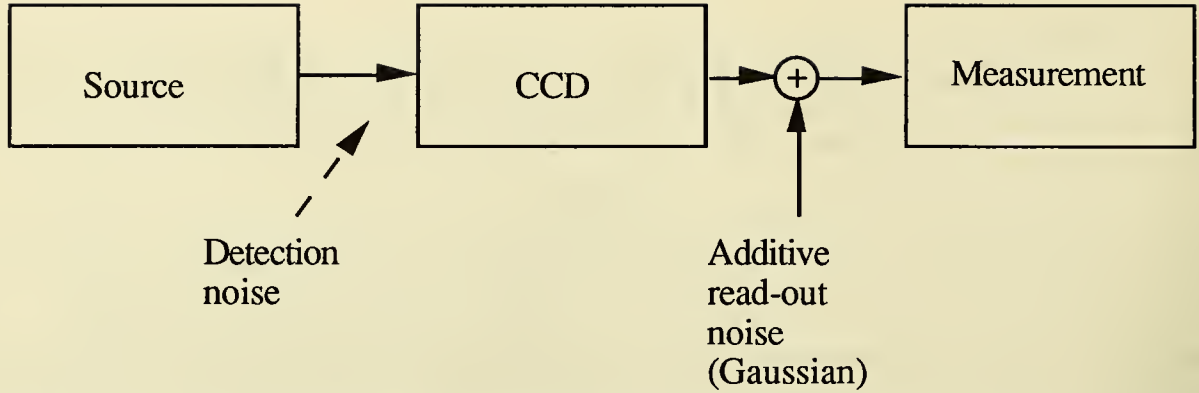
Fig. 2: Model including a Poisson distributed photon source, with photons being received by a CCD camera which preserves the Poisson nature of the data. Electronic readout noise is considered to be additive and Gaussian.

The iterative formula resulting from (10) and (11), obtained by the method of Successive Substitutions[19] is given by

$$a_i^{(k+1)} = K a_i^{(k)} \left[\sum_{j=1}^{B} \left\{ \frac{f_{ji}}{\Delta p_j} \left(\frac{p_j}{\sum_{l=1}^{B} f_{jl} a_l^{(k)}} - 1 \right) \right\} - \frac{1}{\Delta a} \log \left(\frac{a_i^{(k)}}{Q_i} \right) + C \right]^n .$$ (12)

The constant C is arbitrary and it is chosen to insure that the negative terms in (12) do not generate negative image values. The exponent n is an acceleration parameter that, in medical PET reconstructions, can take values $1 \le n \le 3$ with excellent convergence. The speed up of the reconstruction is approximately proportional to n. The constant K is evaluated at each iteration to conserve the number of counts in the image and it is equivalent to calculating the Lagrange multiplier μ of (11).

At the time of this writing, we are proceeding to evaluate the FMAPE algorithm in reconstructions of astronomical images, particularly for the Faint Object Camera in the Hubble Telescope. In that respect it must be pointed out that:

1) the two summations in (12), and also in (6) correspond to convolutions that can be carried out in Fourier domain if the point spread function of the telescope is space invariant.
2) the point spread function, if obtained experimentally, contains stellar background, Poisson noise, fiducial marks, etc., that have to be removed in a judicious way if the MLE or FMAPE reconstructions are to succeed.

6. ALGORITHM FOR POISSON DATA WITH ADDITIVE GAUSSIAN NOISE

The reconstruction of Poisson data from a detector that adds Gaussian noise of a known standard deviation σ can also be treated as a statistical parameter estimation problem. We have developed the following iterative algorithm for that case, which corresponds to the model of Fig. 2.

The target function for a MLE (non-Bayesian) case is the following:

$$L(\mathbf{p}|\mathbf{a}) = \prod_{j=1}^{B} \sum_{k=0}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(k-p_j)^2}{2\sigma^2}} e^{-h_j} \frac{(h_j)^k}{k!}$$ (13)

and the iterative algorithm has the same form as the MLE for the simple Poisson case

$$a_i^{(k+1)} = a_i^{(k)} \sum_{j=1}^{D} \frac{f_{ji} p'_j}{\sum_{l=1}^{B} f_{jl} a_l^{(k)}}$$ (14)

except that instead of using the data p_j we have to use a new set of data p'_j that has to be recomputed at each iteration by evaluating the expression

$$p'_j = \frac{\sum\limits_{k=0}^{\infty} k \, e^{-\frac{(k-p_j)^2}{2\sigma^2}} \frac{\left(h_j\right)^k}{k!}}{\sum\limits_{k=0}^{\infty} e^{-\frac{(k-p_j)^2}{2\sigma^2}} \frac{\left(h_j\right)^k}{k!}}. \tag{15}$$

In the regime of low number of photons and read-out noise being relatively strong, the modified p'_j prevent the MLE algorithm from attempting to estimate the emission parameters too close to the data, providing, in effect, a pixel dependent filtering operation. We believe that this is the first time that this problem has been solved. A practical evaluation of the performance of this algorithm will be undertaken in the near future.

7. CONCLUSIONS

Based on the experience gained in medical tomography, it appears that Maximum Likelihood Estimator methods of image reconstruction, using the concept of feasibility to determine which images are acceptable, and the Bayesian method of regularization with an entropy prior, would be excellent candidates for astronomical image reconstruction. They offer:

1) The ability to use the correct model for the statistical process that generated the data.
2) Space variant point spread functions.
3) Incorporating pixel gain corrections in a statistically correct manner.
4) Unbiased results when the point spread functions is known sufficiently well.
5) Positivity of pixel values is assured.
6) Convergence to a regularized result when using the FMAPE method.
7) Pixel-by-pixel standard deviation which is proportional to the square root of the number of counts in a region.
8) Speed-up of convergence with stability (FMAPE).
9) The FMAPE algorithm does not require the computation of large exponentials, a common problem in entropy prior formulations.

The algorithms discussed are CPU intensive. A medical tomographic image of 128 x 128 pixels, starting from 320 x 64 data points, with space variant point spread function, takes ~ 2 minutes per iteration in a relatively fast Sun4/110 workstation. The very badly posed ET problem requires 25 to 40 iterations for acceptability of a solution, depending on the number of counts in the data set (1 to 3 Million, typically). This is ~ 1/2 the time that would be needed for a conjugate gradient type of algorithm and it does not seem prohibitive for a project of the size of the Hubble Telescope. We have recently evaluated an Intel iPSC/860 hypercube 8-processor system as a possible reconstructor for ET and found that one iteration of the above problem could be obtained in 1.2 second. The cost of the system is in the order of $200,000 and it is likely to drop in the future. Problems with space invariant point spread functions can use Fourier transforms for the convolutions implied by the summations of (6), resulting in much lower computation times.

8. ACKNOWLEDGMENTS

9. BIBLIOGRAPHY

1. H.C. Andrews and B.R. Hunt, *Diginal Image Restoration*, Prentice-Hall, Englewood Cliff, NJ (1977).

2. L.A. Shepp and Y. Vardi, "Maximum likelihood reconstruction for emission tomography," IEEE Trans. Med. Imaging, MI-1, (2), 113-121 (1982).

3. A.D. Dempster, N.M. Laird and D.B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," J. Royal Stat. Soc., B39, 1-37 (1977).

4. J. Llacer, E. Veklerov and J. Nunez, "Statistically based image reconstruction for emission tomography," Interl. J. of Imaging Systems and Technology, 1, 132-148 (1989).

5. L.B. Lucy, "An iterative technique for the rectification of observed distributions," The Astronomical Journal, 79, No. 6, 745-754 (1974).

6. D.G. Politte, "Image improvements in Positron emission tomography due to measuring differential time-of-flight and using maximum likelihood estimation," IEEE Trans. Nuclear Sci., 37, No. 2, 737-742 (1990).

7. J. Llacer and A. Bajamonde, "Characteristics of feasible images obtained from real PET data by MLE, Bayesian and Sieve methods," pres. in SPIE 1990 Intl. Symp. on Optical and Optoelectronic Applied Science and Engineering, San Diego, July 1990, to be publ. Conf. Proc. LBL-29150.

8. E. Levitan and G.T. Herman, "A maximum *a posteriori* expectation maximization algorithm for image reconstruction in emission tomography," Med. Image Processing Group, Dept. Radiology, U. Pennsylvania, Tech. Rep. MIPG115 (1986).

9. J. Nunez and J. Llacer, "A fast Bayesian reconstruction algorithm for emission tomography with entropy prior converging to feasible images," IEEE Trans. Med. Imaging, 9, No. 2, 159-171 (1990).

10. J. Nunez and J. Llacer, "Bayesian image reconstruction in Astronomy," pres. in European Regional Astronomical Meeting, La Laguna (Canarias), July 1989, to be published as a special issue of Astrophysics and Space Science.

11. D.L. Snyder, M.I. Miller, L.J. Thomas and D.G. Politte, "Noise and edge artifacts in maximum likelihood reconstructions for emission tomography," IEEE Trans. Med. Imaging, MI-6, No. 3, 228-238 (1987).

12. E. Veklerov and J. Llacer, "Stopping rule for the MLE algorithm based on statistical hypothesis testing", IEEE Trans. Med. Imaging, MI-6, No. 4, 313-319 (1987).

13. J. Llacer and E. Veklerov, "Feasible images and practical stopping rules in iterative image reconstruction," IEEE Trans. Med. Imaging, 8, No. 2, 186-193 (1989).

14. T. Hebert, R. Leahy and M. Singh, "Fast MLE for SPECT using an intermediate polar representation and a stopping criterion," IEEE Trans. Nucl. Sci., NS-35, 615-619 (1987).

15. J. Llacer, "On the validity of hypothesis testing for feasibility of image reconstructions," IEEE Trans. Med. Imaging, 9, No. 2, 226-230 (1990).

16. E. Veklerov and J. Llacer, "The feasibility of images reconstructed with the method of sieves," IEEE Trans. Nucl. Sci., NS-37, No. 2, 835-841 (1990).

17. L.A. Shepp and R.J. Vanderbei, "New insights into emission tomography via linear programming," NATO Adv. Study Institute Interl. Conf. on the Formation, Processing and Formation of Medical Images, (1988) to be publ.

18. J. Llacer, E. Veklerov and J. Nunez, "The concept of causality (feasibility) in image reconstruction," NATO Advanced Study Institute Intcrl. Conf. on the Formation, Processing and Formation of Medical Images, (1988) to be published.

19. F.B. Hildebrand, *Introduction to Numerical Analysis*, John Wiley & Sons (1974).

On the Problem of Extremely Narrow, Color and Position Dependent, Point Spread Functions

Alan Penny

Rutherford Appleton Laboratory

Introduction

This talk is concerned with the problems that the special nature of the HST point spread function (PSF) will cause in the deconvolution of HST images. The discussion will concentrate on the use of the Planetary Camera mode of the WF/PC.

In its present state, the HST PSF has three main features that are not normally seen in the PSFs of ground-based telescopes, and thus are not allowed for in present deconvolution techniques. The features are:-

- The shape of the PSF is dependent on position in the field-of-view and of the color of the object being observed.

- The PSF has structure on the scale of 0.01 arcseconds over the whole of the area covered (up to 2 arcseconds diameter).

- The PSF has extremely wide wings with respect to size of the central core.

For deconvolution to work, any algorithm must take these features into account.

Futhermore, it will be necessary to determine the details of these features accurately so that the deconvolving algorithm uses the right input PSF. The ways of determining the parameters of these features form the body of this talk.

PSF Position and Color Dependence

This is not completely unknown in ground-based telescopes. Position dependence of the PSF in a ground-based telescope can come from telescope aberrations, from non-orthogonality of the detector to the telescope axis, and from non-planarity of the detector surface. In conditions of good seeing the position-dependence can become marked.

Work on this problem has shown how such a feature can be analysed. In the course of a program of accurate CCD crowded-field stellar photometry, I have developed a software suite (Starman - see Penny and Dickens (1986)) which has been used in an investigation of the abundance spread in globular clusters (see Penny et al., 1989). As only stars are dealt with, it is reasonably simple both to determine the PSF as a function of position and source color, and then to use this determination in the analysis of the star magnitudes.

It does lead to a considerable increase in computing needs as compared with more simple algorithms. This is both in the PSF determination, which in effect has to be done separately for small sub-regions of the image, in terms of a range of star colors, and in the dealing with overlapping star images which, with different colors, have different PSFs. But it does lead to a more accurate PSF, which results in more accurate photometry.

So in a similar fashion, the HST PSF position and color dependence could be determined by measuring the PSFs of stars in different locations and with different colors. To save time, this would presumably be done in a star cluster field which will have many bright stars with differing colors in differing locations. However in such a field the analysis software will, like Starman, need to be able to disentangle the necessarily overlapping images.

Small-scale Structure in the PSF

The small-scale structure in the PSF will cause a major problem in the determination of the PSF. There are two parts to this.

A. In the Central Core

The central peak has all the sharpness originally intended for the HST. An analysis of the PC PSF image supplied by STScI shows that the central core is so narrow that the intensity falls dramatically from the central pixel to the six adjacent pixels. An analysis using Starman shows that the intrinsic PSF, before the averaging and summing into the 0.04 arcsecond pixels of the PC, has a profile such the intensity falls to one half in a distance of 0.0345 pixels (0.014 arcseconds).

Thus the observed, pixilated, star profile will change significantly if the star center, with respect to the pixel centres, moves by 0.1 pixels. So to determine the actual PSF accurately, star images will need to be obtained at the one hundred positions of a grid, made up of 0.1 pixel steps of star centre to pixel center distance. This will have to be done at different locations in the field-of-view as the PSF changes with position, and for stars of different colors as color too affects the PSF. A truly substantial observing and data analysis program.

B. In the Wide Wings

Investigation of the PC PSF image also shows that even far out, certainly at 0.5 arcsec from the centre, the PSF shows extremely fine small-scale structure. Adjacent pixels can have intensities varying by a factor of three. This again means that the actual PSF in the wings, before the pixel averaging, has structure on scales small with respect to the pixels. This shows that the determination work for the core will have to extend out to 1 arcsecond from the center, resulting in a PSF determined for a 500x500 grid, with the 50x50 grid of pixels out to 1 arcsecond radius sub-divided by the 10x10 sub-pixel sampling.

Furthermore, as these wings have intensities only a few percent of the peak intensity, many exposures will have to be coadded to give sufficient signal-to-noise in the wings, whilst not saturating the core.

The Size of the Wide Wings

The wide wings mean that in a calibrating exposure to a star cluster to determine the PSF, it will be common for the wings of stars to overlap. Thus conventional PSF fitting software, which commonly takes a grid of points containing the PSF, will have to take a account of neighbours. A simple calculation shows that having a 500x500 point PSF will lead to prohibitive computing needs in the multiple star fits of the PSF determination. It is likely that modifications of the algorithms will be needed, to subtract the wide wings in a truncated iterative manner before the detailed fitting.

The Combination of Observed and Calculated PSFs

As the previous three sections have shown, the determination of the PSF will be an extremely extensive matter. It will clearly be advantageous if direct observations can be backed up by theoretical calculations.

An accurate determination of the HST mirror shapes will come from investigations of out-of-focus images as proposed by Lasenby elsewhere in this workshop, from ray-tracing techniques, and from a knowledge of the error in the null corrector. This should mean in theory that the PSF for any position and wavelength can be calculated. If this can in fact be done, and a comparison with observations shows that the desired accuracy is reached, then the number of such observations needed would be drastically reduced. Only a small grid of observations would be needed to check on the theory.

A Reminder that the other features that affect the PSF

In all this discussion of determining the PSF, there are other substantial problems which will make the matter complicated. These are the existing, well-known problems in the detectors themselves, which make the detectors not quite perfect two-dimensional detectors.

A short summary of these are:- For the WF/PC, the hysteresis, the problem of accurately determining the flat fields, the steps in the D/A converter, and the charge transfer effects, and cosmic rays; For the FOC, the geometric stability, the flat fields, the non-linearity, and charge pulling.

All of these are planned to be investigated during the Science Verification, but until this occurs and these problems are understood the PSF determinations will be suspect. It will also be necessary to understand how accurately these matters have to be known for PSF determination needs.

Some PSF Determination Questions

This complicated PSF will lead to extensive effort both in determining it and in using it to deconvolve and to measure images. For confidence in the science done it will be necessary to know what accuracy in the PSF has been achieved. The planning of the observations will need to know what accuracy will result from a feasible amount of effort. Thus basic questions result.

- What accuracy is needed in the PSF to permit good science?

- How many observations and reduction analysis are needed to determine to PSF to this accuracy?

- How much use will a complicated PSF be in the deconvolution or measurement of an image, given limitations of data analysis computing resources?

References

Penny,A.J., and Dickens, R.J., 1986. M.N.R.A.S., 220,845.

Penny,A.J., Buttress,J., and Griffiths,W.K., 1989. A.S.P.Conference Series, Vol. 8., p305. 'CCDs in Astronomy', Ed. G.H.Jacoby

Remarks on Hard Modeling vs. Image Processing, Circumstellar Disks, and Model Validation

Donald Geman
Department of Mathematics and Statistics
University of Massachusetts at Amherst
Amherst, Massachusetts 01003

January, 1990

NOTE: *This contribution is a slightly revised version of the introductory section of "Statistical inference on the shape of circumstellar disks based on HST observations," co-authored with P. Dupuis, J. Horowitz, and G. Reynolds (address above) and S. Strom, Five College Astronomy Department, University of Massachusetts. The modifications are intended to emphasize my remarks at the Workshop about "off-the-shelf" solutions and performance evaluation.*

1   Hard Modeling vs. Image Processing

Despite the advantages of placing a telescope outside the earth's atmosphere, the images taken with the Wide Field Planetary Camera (WF/PC) aboard the HST are still considerably degraded, mainly due to the unexpected, severe blurring resulting from the infamous aberration in the mirror(s), but also due to the usual limitations imposed by photon-limited data and other effects of the image formation system. As a result, the amount of information from HST data that is readily accessible (e.g., visually evident) may be limited. Instead, in order to estimate quantities of interest, some form of "processing" will often be necessary, perhaps involving image restoration and/or statistical inference.

The standard approach to image restoration is non-dedicated and nonparametric: except for specific knowledge of the image formation process, it incorporates only *generic* assumptions, for example constraints on the positivity, smoothness, or entropy of the brightness pattern. Many examples of such techniques were advocated at the Workshop, including those based on pseudo- inverses, maximum entropy, maximum likelihood (e.g., Lucy-Richardson iteration), "regularized" maximum likelihood, and Bayesian inference with "prior" and "posterior" distributions. We shall refer to these approaches as "soft modeling." Usually, they

result in a single "restored image," deemed to capture the original brightness pattern without degrading effects, or at least suppress noise and enhance resolution. Since there is no "scene model" *per se*, there is no systematic method for *directly* estimating physical scene characteristics, let alone assessing their precision. Still, in the absence of additional information, little else can be done. For example, it is unlikely that the brightness pattern of entire galaxies, or other such relatively unstructured objects, can be analytically (or otherwise economically) encoded.

In contrast, in some astrophysical problems, physical evidence or arguments may supply extra information, for example mathematical constraints or even precise, global hypotheses, which permit the underlying source distribution to be represented by a parametric scene model. In our case, for example, the scene model is based on geometric and photometric properties of the pattern of light scattered by the material in a circumstellar disk. This approach is then problem-specific and model-based; we have dubbed it "hard modeling" to emphasize the existence of an explicit model of the physical situation. The information sought is then *parametrically* encoded; for example, there may be parameters corresponding to the disk radius, orientation, and brightness fall-off. The direct results of this analysis are then *parameter estimates* based on the data rather than a restored image. In adopting his approach, we were motivated partly by our failure to get accurate restorations from simulated HST data with standard image recovery techniques, but perhaps more so by our conviction that it is more promising for the disk problem.

In either case, it is unlikely that many of the astronomical problems of interest have off-the-shelf solutions, and no methodology enjoys "axiomatic privilege." On the contrary, each should be validated in a convincing manner, *which requires more than simply displaying successful experiments.* In the case of soft modeling, this could be done by guaranteeing the photometric accuracy of restorations relative to a series of templates, i.e., a prototype class of images. More specifically, if the stochastic inverse problem is reformulated as an optimization problem, which is usually the case, than one might *prove* (in simulations) that the original (known) source distribution is indeed an extremal state, or close to it. This is likely to be difficult; see [1] for a primitive effort in this direction, and [4] for additional remarks on this issue. On the other hand, in the case of hard modeling, this involves deriving constraints on the precision of the parameter estimates (i.e., confidence statements), perhaps devising formal statistical tests of competing hypotheses, and certainly diagnosing the model based on the residuals, since this approach will automatically result in a "pretty picture" (namely the one corresponding to the estimated parameters), regardless of the source distribution. This will be amplified in §4.

2   Circumstellar Disks and Planetary Systems

At present, there are no known examples of planetary systems other than our solar system, in which the orbits of the planets all (nearly) lie in the equatorial plane of the sun. It is

conjectured that the sun was once surrounded by a disk composed of distributed dust and gas, out of which the planets agglomerated, the residual material blown away by high energy winds along the polar axis of the sun.

More generally, astronomers believe that young stars throughout the universe are surrounded by "circumstellar disks" and that these are the environment in which planetary systems are developed. Apparently, this brief episode of stellar evolution is part of a broader scenario, only loosely understood, but thought to begin when a cold, rotating protostellar core condenses inside a large molecular cloud to form a star-disk system. Eventually, the star enters the "main sequence" (i.e., hydrogen-burning) stage, possibly accompanied by a planetary complex and other disk remnants.

To assess the accuracy of this theory, astronomers wish to use the Hubble Space Telescope (HST) to image the environment of nearby, *pre-main sequence* stars (both single and binary) in order to determine the frequency of occurrence of such disks and to characterize their physical characteristics, such as size, shape, mass, and surface brightness, during the period from stellar birth until the disks become "optically thin." (Apparently, direct imaging of planets is beyond current technology.) One objective is then to "... learn for the first time *from direct observations* whether our planetary system is likely to represent a common and expected result of star formation, or whether formation of our Solar System required a fortuitous combination of circumstances reproduced only rarely in nature" ([5]). In fact, in the former case, "all single stars and well separated binaries cooler than [spectral type] F0 might be expected to have planetary systems. We might therefore expect planetary systems to occur around one star in three, some 10^{12} planets in the whole Galaxy" ([3, p.307]).

Aside from our own solar system, the current direct evidence for the existence of circumstellar disks consists of one that was observed in 1984 around β Pictoris. An "extended object" observed by S. Strom and co-workers was originally thought to be a disk, but this interpretation has recently been called into question (Strom, personal communication). There is also spectral evidence for the presence of disks; for example, the starlight is absorbed by dust particles in the disk and re-radiated at longer wavelengths, resulting in significant energy at infrared and other frequencies. Finally, there is indirect evidence for large planets derived from perturbations in stellar trajectories and velocities.

3   Formulation as a Problem in Statistical Inference

Let $I_\theta(x)$, $x \in \mathbf{R}^2$, denote a *parametric family of intensity models*, where $\mathbf{R}^2$ represents the "plane of the sky", θ is a vector of parameters which incorporates various geometric and photometric properties of the star-disk system, and I_θ, the scene model, is the intensity distribution of the light coming from the star-disk system. For example, in one model for a circumstellar disk, we take $\theta = (\theta_1, \theta_2, \theta_3, \theta_4)$, where $\theta_1 = $ orientation parameter, $\theta_2 = $ disk radius, $\theta_3 = $ flaring parameter, and $\theta_4 = $ intensity of the central star. Thus each such θ corresponds to a specific disk model. In Figure 1 we show one such example which
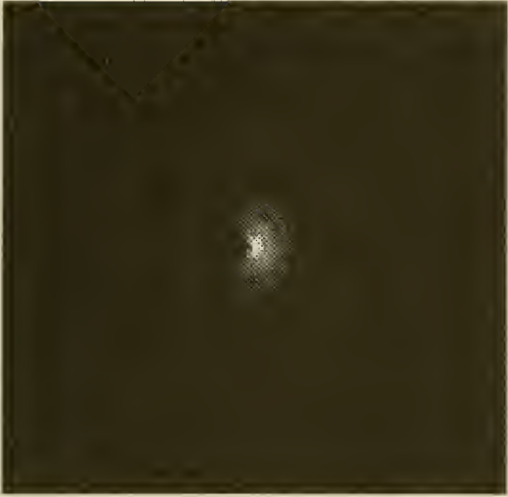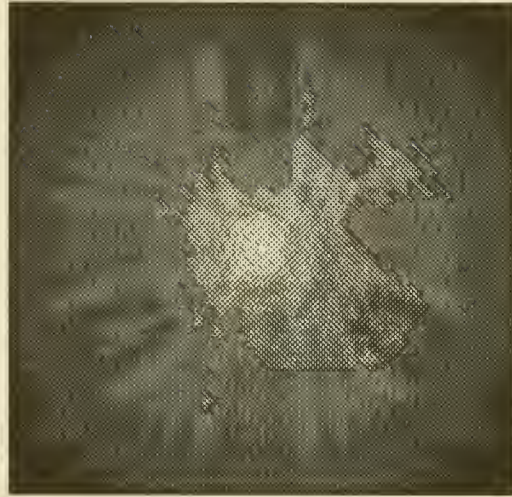
Figure 1: Disk.

Figure 2: Hubble psf.

incorporates a simple scattering model; due to the dominance of the central star, the data has been scaled to make the disk visible.

Obviously, it is necessary to have an accurate model of the image formation process. The optical system will blur the scene (diffraction effects) and the physics of photon detection implies there is inherent random noise in the acquisition. This alone may place non-trivial constraints on the amount of information that can be obtained, as emphasized by several of the Workshop attendees. In addition, and unfortunately, the point spread function (PSF) for the WF/PC has significant mass over a radius on the order of one arc-second (see Figure 2), resulting in severe distortion. Finally, exposure times are limited by the instability of the PSF, as well as "bleeding." We model the image formation process using the semiclassical model ([2, ch.9]) of photodetection, which means that quantum effects are accounted for only at the detection end of the system. In summary, *after* passing through the optical system and impinging on the photosensitive detector surface, the probability distribution for the observed counts is

$$P(n_1, n_2, ..., n_J) = \prod_{j=1}^{J} e^{-\tau\lambda_j(\theta)}(\tau\lambda_j(\theta))^{n_j}/n_j! \qquad (3.1)$$

where n_j, $j = 1, ..., J$, is the photon count in the j'th detector (i.e., pixel), τ represents the exposure time, and $\lambda_j(\theta)$ is the intensity of light falling on the j'th detector after incorporating the effect of the PSF. Various other effects are not incorporated here, such as saturation, quantum efficiency, and read-out noise.

The problem is to estimate (or test hypotheses about) θ based on the observations $\{n_j\}$. Once the model I_θ is specified, we can (in principle) compute maximum likelihood estimates (MLEs) $\hat{\theta}$ of the parameters based on (3.1). (Other methods of estimation are possible, but we will stick to MLE.) Aside from obtaining the actual parameter estimates (there are some computational problems), the primary objectives are then to obtain confidance statements

and determine goodness of fit; more specifically, to assess the accuracy of estimation given that the scene actually satisfies the proposed model (or "null hypothesis"), and to determine whether the disk model actually fits the physical phenomena being observed, including testing this model against plausible alternative hypotheses.

4   Model Validation

Concerning parameter estimation, the most important problem is to determine the estimation error for realistic values of τ and J, meaning for feasible exposure times and the actual resolution of the WF/PC. This problem will necessitate formulating an appropriate metric d for measuring the distance $d(\theta^*, \hat{\theta})$ between the true parameter θ^* and its estimate $\hat{\theta} = \hat{\theta}(\tau, J)$. Then the function

$$E(\tau, J; \epsilon) = \max_\theta P_{\theta, \tau, J}(d(\theta, \hat{\theta}) > \epsilon)$$

describes the quality of inference that can be expected with error ϵ for the design parameters τ and J. Whereas E is difficult to study analytically (for fixed τ and J), it appears that it can be determined from Monte Carlo simulations and that "order of magnitude" estimates of θ^* that are valid with *very high probability* are of interest in astronomy. Thus, for example, an estimate of the disk radius to within thirty (or even fifty) percent with an error probability of only a few percent might be desirable. We have done some empirical estimates of $E(\tau, J; \epsilon)$ by tracking the accuracy of the parameter estimates as a function of exposure time, with J fixed. The results are preliminary, but suggest that, given the existence of a star-disk system, accurate estimates of its properties are possible well below saturation levels, and that we can in fact "trust" certain conclusions which are not already visually apparent in the data.

We have also investigated the *consistency* of the MLE's; in particular, we show, using a large deviation argument, that under certain "identifiability" assumptions, the parameter estimates converge to the true values as the exposure time τ grows to infinity; here J is fixed.

The goodness of fit problem is important because, as mentioned above, one can always estimate the parameters from (3.1) and generate a picture $I_{\hat{\theta}}$ of a disk, no matter what the *actual* shape of the light source. Plausible *alternative hypotheses* must be formulated and evaluated against the data. Our simulations will compare the primary hypothesis, namely that of a star-disk system consisting of a point source star embedded within a flaring disk, against several alternatives, such as a binary star and a "gap" model, in which there is a space around the star containing no disk material.

At this preliminary stage, we have done no more than "look" at the residual images obtained by fitting the star-disk model to data generated by the competing scenarios. More precise diagnostics will eventually be necessary, including tests on the residuals to help confirm model fit or lack of fit, especially tests which are sensitive to data generated by

a binary star or by a disk with a gap as opposed to a "full" disk. Some tests for planar white noise are in the spatial statistics literature, and there are results about "deviates" from the theory of *linear* models, but none of these are really designed for the case at hand. In addition, we intend to utilize likelihood ratios for testing the primary hypothesis against both single and composite alternative hypothesis, although rigorous results may be elusive.

It is doubtful that any statistical test will be sufficiently powerful against a sufficiently rich class of alternative hypotheses to conclude that the data is *uniquely* explained by a primary hypothesis. Perhaps the best one can do is to determine that a primary hypothesis is consistent with the data, and get useful information about the corresponding star-disk system, for which preliminary results are encouraging.

REFERENCES

1. Geman, D. and Reynolds, G., "Constrained Restoration and the Recovery of Discontinuities," Technical Report, University of Massachusetts, 1990.

2. Goodman, J.W., *Statistical Optics*, Wiley, New York, 1984.

3. Kitchin, C.R., *Stars, Nebulae, and the Interstellar Medium*, Adam Hilger, 1987.

4. Snyder, D.L., "Modifications of the Lucy-Richardson Iteration for Restoring Hubble Space Telescope Imagery," Technical Report, Electronic Systems and Signal Processing Laboratory, Washington University, 1990.

5. Strom, S. et al, "The Formation and Evolution of Solar Nebulae Surrounding Pre-Main Sequence Stars," proposal submitted to the Space Telescope Science Institute, 1988.

RESTORATION WITH INCREASED SAMPLING - IMAGES AND SPECTRA

L.B. Lucy

European Southern Observatory

Introduction

With respect to both images and spectra, the desire for wide coverage coupled with the limited array size of digitized detectors commonly results in a poor sampling of the instrumental response function. This led to work at ESO (Lucy and Baade 1989) in which a rebinning option was incorporated into a deconvolution package. A prime motivation for this development was in fact the HST since, according to specifications, it was expected to poorly sample the PSF in the f/48 imaging modes.

A further reason for rebinning is the displeasing appearance of modern astronomical images. This results partly from the coarse gridding imposed by the detector but most especially from pixel-to-pixel noise, since this destroys continuity and thus makes the eye acutely aware of the grid. In addition to this aesthetic consideration, one can well imagine that this prominence of the grid is deleterious for morphological and classification investigations of images.

Images

Deconvolution with simultaneous resampling can be effected by incorporating the gridding operation into the mathematics of image formation (Lucy and Baade 1989). This yields an integral equation that relates $\psi(\xi)$, the intensity distribution on the sky, to $\phi(x)$, the discontinuous function giving the intensity distribution in the pixelated focal plane. Application of a deconvolution technique to this modified equation of image formation then allows $\psi(\xi)$ to be sampled on a finer grid than that of the detector while retaining the observational data $\tilde{\phi}(x)$ in their original form. Accordingly, there is no degradation of the observational data.

This ESO code has been used to explore the restoration possibilities for HST images using a simple model for the flawed PSF due to the residual spherical aberration. Results for a simulated image of a close double star have been published (Baade and Lucy 1990) and so need not be displayed here. It suffices to remark that the results are distinctly encouraging and surely demonstrate that image restoration will play a vital rôle in allowing some at least of the scientific goals of the HST to be achieved. In addition, the published example nicely illustrates the positive contribution of rebinning. Because of the increased sampling and the near absence of pixel-to-pixel noise in the restored image, one is barely aware of the grid.

Error analysis

The restored image given by iterative deconvolution (Lucy 1974) can be written as

$$\psi(\xi) \;=\; \int \tilde\phi(x)\, Q\,(\xi|x)\, dx, \tag{1}$$

where $Qd\xi$ is the derived estimate of the probability that a photon recorded at x would with a perfect instrument have been recorded in $d\xi$ at ξ. Because the photons comprising the observed image $\tilde\phi(x)$ are thus redistributed exactly as specified by the pdf Q, we see that $\psi(\xi)$ is the estimated expectation value of the intensity at ξ. Accordingly, we can distinguish two sources of uncertainty in the restored image: First, the uncertainty in $\psi(\xi)$ due to sampling errors in $\tilde\phi(x)$. Second, the uncertainty due to statistical fluctuations about $\psi(\xi)$ in consequence of the finite number of photons in the observed image.

Errors of the first type can be quantified by noting that equn. (1) is a linear mapping and thus allows a conventional propagation-of-error analysis (Lucy 1974, Sect. IV).

Errors of the second type can be quantified by using Monte Carlo sampling to redistribute photons according to $Q(\xi|x)$. A sequence of such Monte Carlo images, each of which is a feasible variant of the restored image, could (following Skilling) be made into a video, thereby allowing the reliability of features in the restoration to be assessed. Note that uncertainties thus explored are those due to the restoration - $\tilde\phi(x)$ is here assumed to be exact.

Limited resolution restoration

When deconvolving an astronomical image, we are attempting to recover the true intensity distribution on the sky - i.e., to recover what we would observe if the aperture and S/N both $\to \infty$. By posing this problem, we are challenging our deconvolution procedures to reproduce generalized functions - e.g., to represent stars as delta functions - and this requires accurate recovery of Fourier amplitudes for wavenumbers $\to \infty$. Because this is impossible, deconvolved images contain artefacts having the character of Gibbs oscillations. Clearly, if we limit the resolution of the deconvolved image, such artefacts should be reduced in amplitude.

In the context of the degraded HST with its PSF of startling complexity, we might be well advised to limit our restoration ambitions to the resolution achievable with a perfect HST. We could call this restoration to the status quo ante or restoration to nominal. In this case, we would not use $P_d(x|\xi)$, the observed or "dirty" PSF when deconvolving HST images but the PSF P_c, gives by

$$P_d(x|\xi) = \int P_c(x|x') P_i(x'|\xi)\, dx', \tag{2}$$

where P_i denotes the PSF of a perfect HST. In the non-HST context, P_i is the PSF of the finite aperture telescope whose ideal performance is the upper limit to our restoration ambitions.

A simple, 1-D example of limited resolution restoration is given in Fig. 1. Here the seeing-broadened image of a star on a flat background is deconvolved both with (B) and without (A) a limit on the resolution. The seeing gives a Gaussian PSF with $\sigma_d = 1$, the pixel size $= 1$, and the ideal PSF has $\sigma_i = 1/\sqrt{2}$. With no limit on the resolution, the deconvolved stellar profile (A) eats significantly into the background, resulting in slow convergence of an integration for the total intensity. In contrast, when the resolution is limited (B), the artefact has a markedly decreased amplitude and also a decreased angular scale.

In the simple case of stars on a uniform background, this artefact can be avoided by subtracting the background. An algorithm that does this for noisy backgrounds and yet preserves non-negativity has been described (Baade and Lucy 1989).

A further discussion of limited resolution restoration is given by Snyder (this volume).

Spectra

As discussed by Gilliland (this volume), spectra obtained with the large science aperture of the GHRS will suffer from a degraded response function in direct consequence of the degraded PSF of images. Accordingly, we face a problem of restoration to the status quo ante - i.e., to recover the spectrum that would have been observed with the anticipated response function.

In attacking this problem, we must be concerned about artefacts. With non-linear restoration algorithms, an artefact similar to that illustrated in Fig. 1 arises when the spectrum contains a narrow emission line on a continuum. In the spectral case, we must be less tolerant of such artefacts because of the greater likelihood of spurious astrophysical theories being thereby prompted. In view of this, the CLEAN approach of Högbom (1974) has been investigated for the restoration of HST spectra.

Let $\tilde{F}_\lambda$ denote the observed spectrum, $\tilde{P}_\lambda$ the flawed response function, and P_λ the ideal response function. The steps in obtaining the restored spectrum F_λ are then as follows:

(1) Initialization: The residual dirty spectrum $R_\lambda = \tilde{F}_\lambda$, and the current estimate of the restored (cleaned) spectrum $F_\lambda = 0$.

(2) $\tilde{P}_\lambda$ is cross-correlated with R_λ and well separated peaks located.

(3) At the position of the k th peak, $\alpha_k \tilde{P}_\lambda$ is subtracted from R_λ and $\alpha_k P_\lambda$ added to F_λ. Here the α_k are chosen so that only a small fraction of each cross-correlation peak is removed at each cycle.

(4) Stop if the process has converged. If not, go to step (2).

Note that in building up the cleaned spectrum F_λ a finer binning may be used than that of the recorded spectrum $\tilde{F}_\lambda$.

Cleaning and resampling experiments with the above procedure have been carried out with the dirty and ideal response functions shown in Fig. 2. Two simple examples are reported here. In the first, a strong unresolved emission line stands on a flat continuum. The simulated dirty spectrum and the infinite S/N ideal spectrum are plotted in Fig. 3. These may be compared with the cleaned and resampled spectrum and the corresponding-ly finely sampled ideal spectrum plotted in Fig. 4. We see that the cleaned spectrum rather successfully reproduces the ideal profile but noise fluctuations have been amplified - this is the penalty for the flawed response function. When the cleaned spectrum (Fig. 3) is compared to the infinite S/N but coarsely sampled ideal spectrum, the resampling option is seen to be highly beneficial.

In Figs. 5 and 6, similar results are shown for a partially resolved absorption line. In this case, cleaning and resampling is successful in recovering the near blackness of the line's centre.

These experiments suggest that this 1-D CLEAN algorithm with resampling option is a promising technique for restoring HST spectra obtained with the large aperture of the GHRS. Application to spectra obtained with ground-based telescopes should also be considered.

Acknowledgements

For information and discussions, I am grateful to H.-M. Adorf, D. Baade and M.Rosa.

References

Baade, D. and Lucy, L.B.: 1990, ESO Messenger No. 61, p. 24.

Baade, D. and Lucy, L.B.: 1989, in Proc. 1st ESO/ST-ECF Data Analysis Workshop, eds. P.J. Grosbøl, F. Murtagh, R.H. Warmels, p. 169.

Högbom, J.A.: 1974, Astron. Astrophys. Suppl. 15, 417.

Lucy, L.B.: 1974, Astron. J. 79, 745.

Lucy, L.B. and Baade, D.: 1989, in Proc. 1st ESO/ST-ECF Data Analysis Workshop, eds. P.J. Grosbøl, F. Murtagh, R.H. Warmels, p. 219.

Fig. 1: Star on flat background.
Restoration with (B) and without
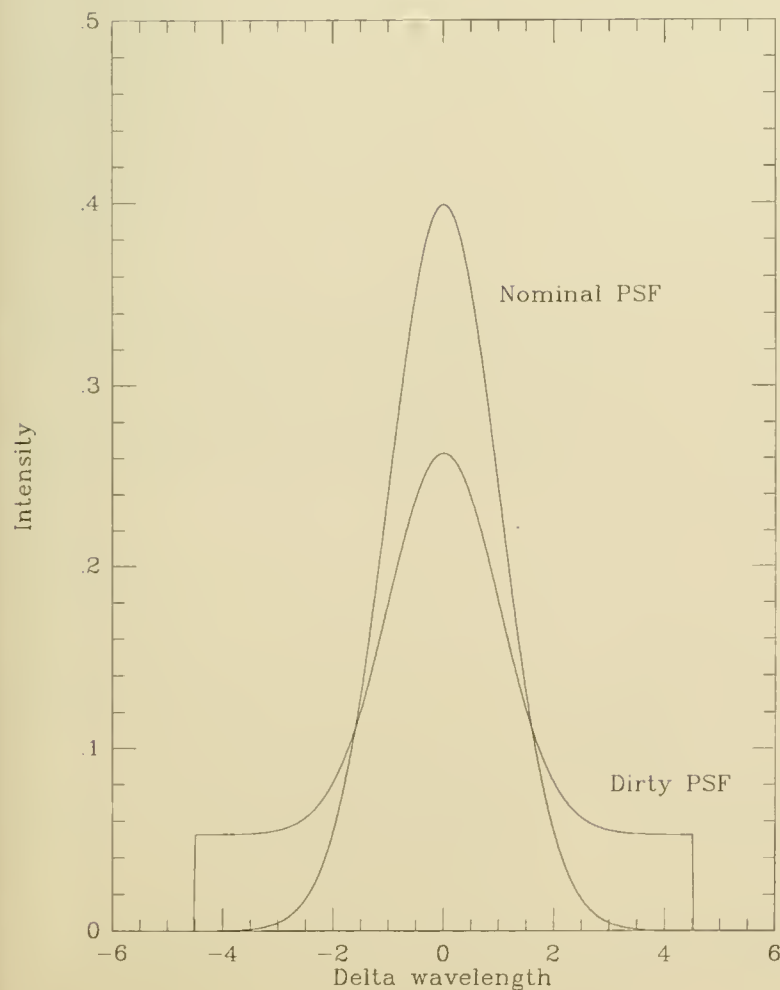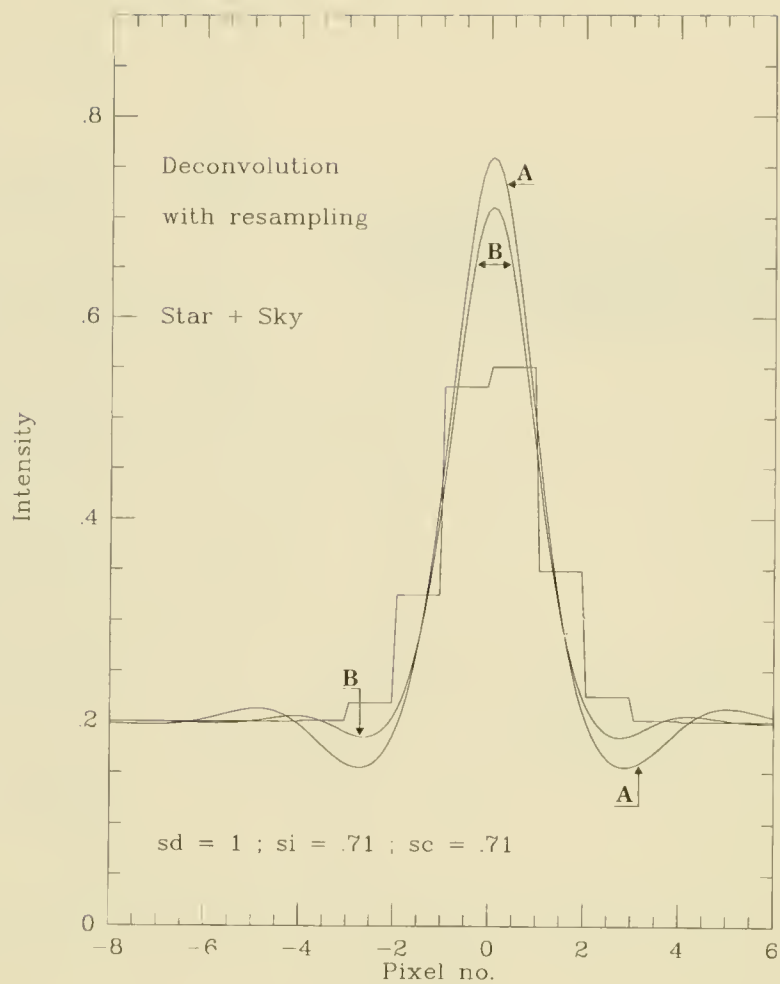(A) limit on requested resolution.

Deconvolution
with resampling

Star + Sky

sd = 1 ; si = .71 ; sc = .71

Fig. 2: Models for flawed and
ideal response function of GHRS
with large science aperture.

Nominal PSF

Dirty PSF

Fig. 3: Emission line on flat continuum. Ideal spectrum with S/N = ∞ also shown.

Dirty spectrum
Poisson noise
Coarse grid

Cleaned & Resampled

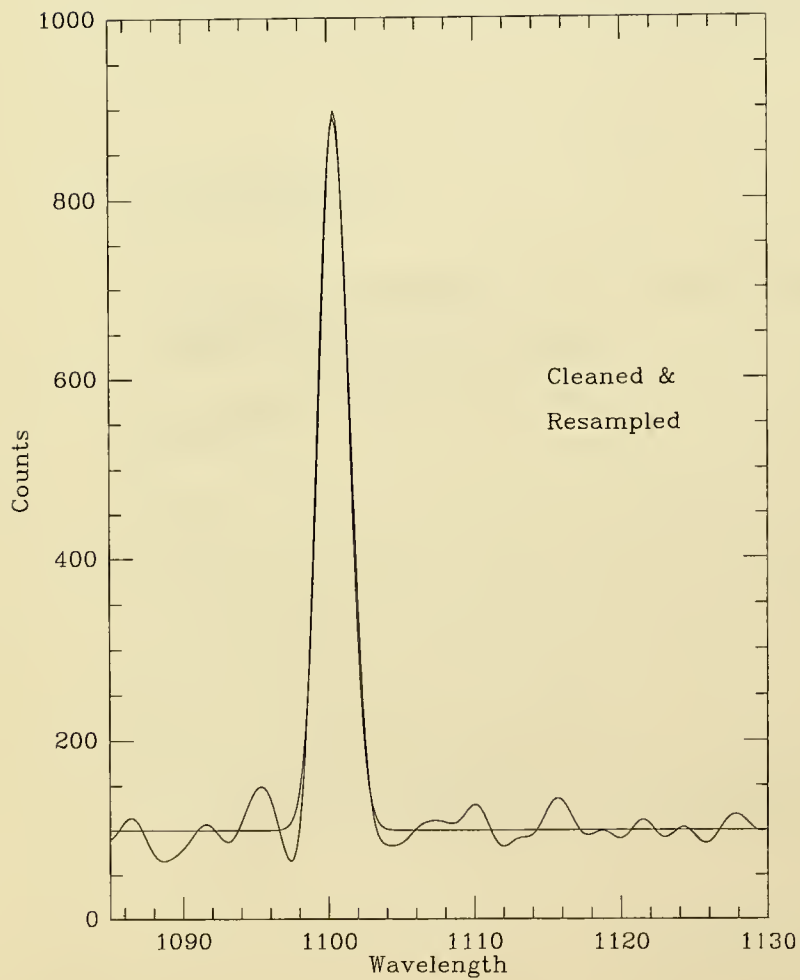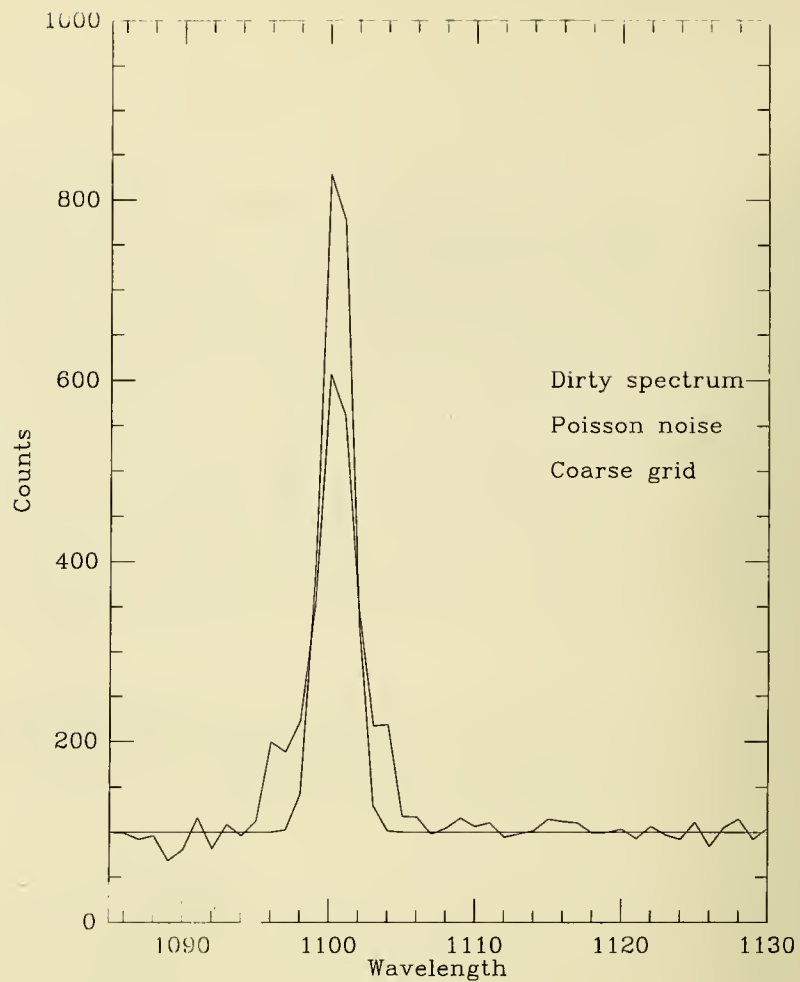Fig. 4: Observed spectrum from Fig. 3 after cleaning and re-sampling. Finely sampled ideal spectrum with S/N = ∞ also shown.

Fig. 5: Flat continuum with
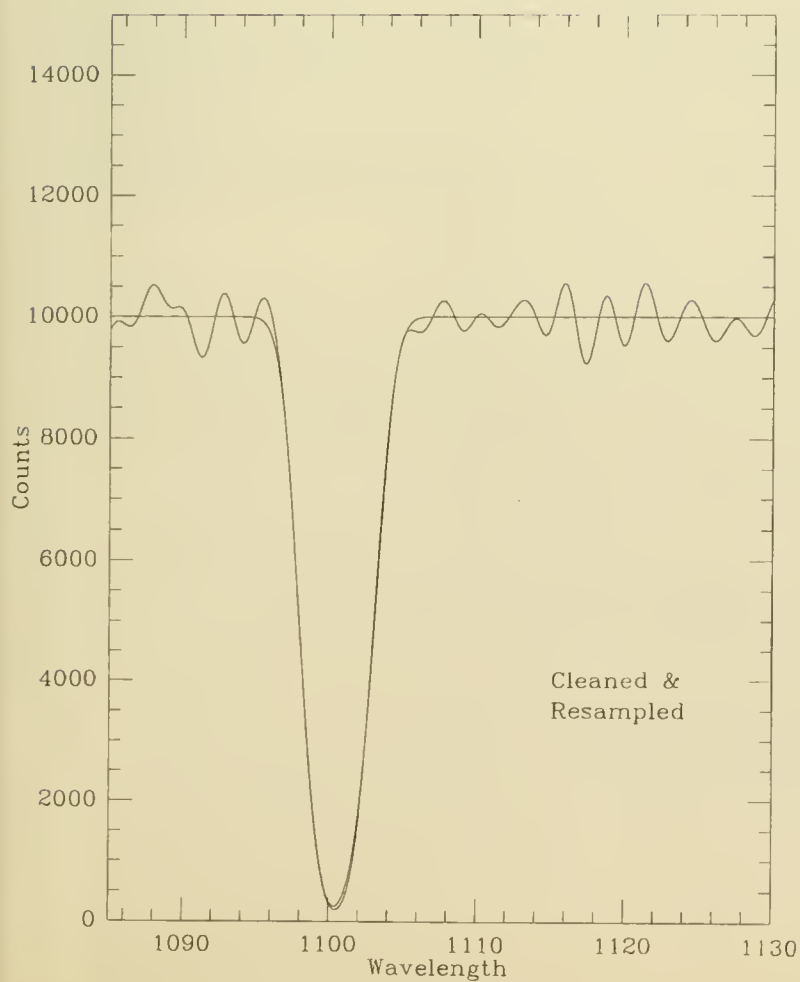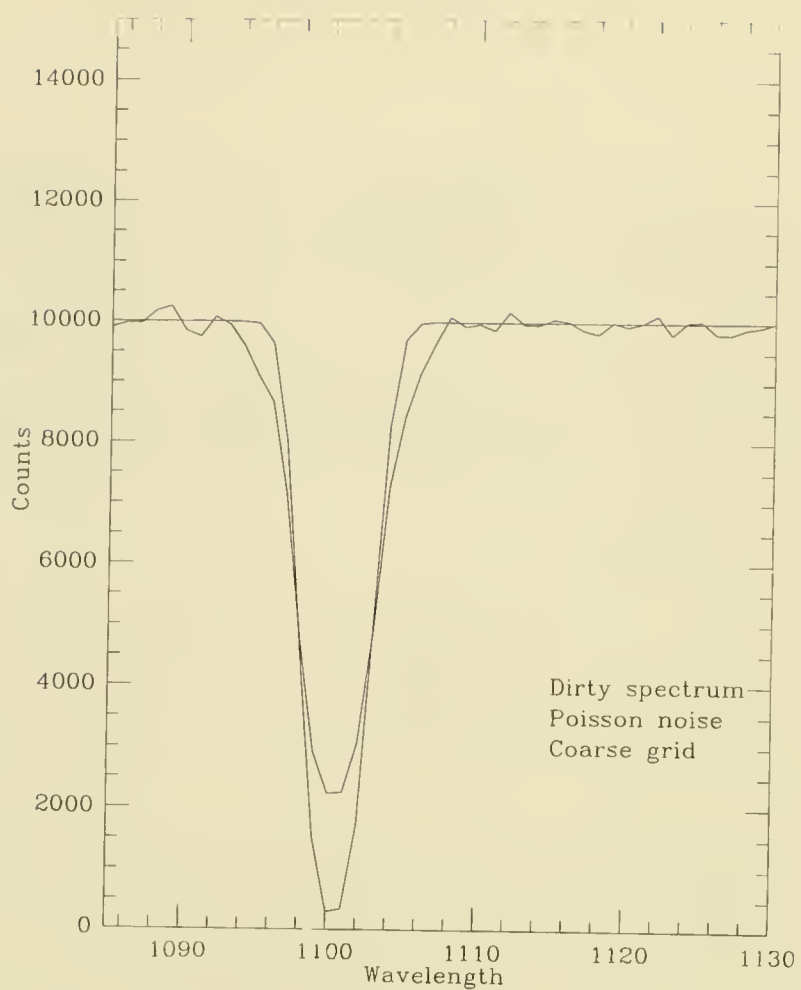absorption line. Ideal spectrum
with S/N = ∞ also shown.

Dirty spectrum
Poisson noise
Coarse grid

Wavelength

Counts

Fig. 6: Observed spectrum from
Fig. 5 after cleaning and re-
sampling. Finely sampled ideal
spectrum with S/N = ∞ also shown.

Cleaned &
Resampled

Counts

Wavelength

87

Deconvolution of Hubble Space Telescope Data: Computer Simulations and Laboratory Experiments

T. Reinheimer and G. Weigelt

Max-Planck-Institut für Radioastronomie
Auf dem Hügel 69, D - 5300 Bonn 1, Fed. Rep. of Germany

1. Introduction

The point spread function (psf) of the HST consists of a central core of about 0.1" diameter and a halo (diameter of several seconds of arc) caused by spherical aberration. At shortest UV wavelengths the psf core breaks into several speckles. Therefore, HST images carry high-resolution information in spite of the spherical aberration.

We have studied the deconvolution of simulated HST data by computer simulations and laboratory experiments. We will discuss our results obtained with the roll deconvolution technique (Lohmann and Weigelt 1979, Müller and Weigelt 1987), inverse filtering, the van-Cittert method (van-Cittert 1931) and the Gerchberg method (Gerchberg 1974). The advantage of the computer and laboratory experiments is that we we can study the dependence of the SNR in the reconstructed image on aberrations and photon noise in the raw image.

2. Computer simulations

We have applied inverse filtering, the van-Cittert method and the Gerchberg method to simulated HST data. The van-Cittert method is an iterative method which calculates the (k+1)th iteration of the reconstruction by the algorithm

$$o_{k+1}(x) = o_k(x) + [i(x) - (o_k(x) * p(x))],$$

where $i(x)$ is the raw image degraded by aberrations and photon noise, $p(x)$ is the point spread function of the telescope, x is the 2-dimensional image plane coordinate, and the asterisk $*$ denotes the convolution operator. As first object estimate one can, for example, use the image deconvolved by inverse filtering.

The iterative Gerchberg method consists of the following image processing steps:

(1) The k th iteration of the object intensity distribution is called $o_k(x)$. As first estimate $o_0(x)$ of the object intensity distribution one can choose the reconstruction obtained by inverse filtering.

(2) Calculation of the Fourier transform $O_k(u)$ of $o_k(x)$.

(3) Calculation of a new estimate

$$O'_k(u) = O_0(u) W(u) + O_k(u) [1-W(u)],$$

where $O_0(u)$ is the object Fourier transform of the first estimate (e.g., obtained by inverse filtering) and $W(u)$ is a weight function. The weight function is large (~ 1) where the modulus of the Fourier transform of the psf has large values and the weight function has small values where the modulus of the Fourier transform of the psf has small values (i.e., around zeros). The above equation describes the following calculation. At coordinates u where $W(u)$ is large (~ 1), the values of $O_k(u)$ are essentially replaced by $O_0(u)$ (Fourier constraints), i.e. the main contribution to $O'_k(u)$ comes from $O_0(u)$. At coordinates where $W(u)$ is very small, the object Fourier transform is not known very well and, therefore, no big changes are made to $O_k(u)$.

(4) Calculation of the inverse Fourier transform $o'_k(x)$ of $O'_k(u)$.

(5) Calculation of a new object estimate $o_{k+1}(x)$. The new estimate $o_{k+1}(x)$ is derived from $o'_k(x)$ by setting $o'_k(x)$ to zero in the region where the object is known to be zero and where $o'_k(x)$ is not ≥ 0 (object constraints). The new object estimate $o_{k+1}(x)$ is the starting point for the next cycle.

   Figure 1 shows a computer simulation of deconvolution by inverse filtering and by the Gerchberg method. The figures show (a) the object of the computer simulation (star cluster), (b) the point spread function, (c) the image obtained by convolving the object with the psf and by simulating photon noise of 50 000 counts per image (about 30 counts per pixel in the brightest pixel), (d) the high-resolution image reconstructed from Fig. 1c by inverse filtering (with interpolation of the object spectrum at coordinates, where the transfer function was very close to zero) and (e) the image reconstructed by the Gerchberg method. Fig. 2 shows a similar computer simulation as Fig. 1. The object is a galaxy and photon noise is 200 000 counts per image.

3. Laboratory Experiments

Figs. 3 and 4 show deconvolution experiments made with laboratory data. The aberration degraded images were produced in a laboratory setup which consisted of the following components:
(a) a lamp for illuminating the laboratory objects (various slides),
(b) optics for simulating the HST,
(c) aberration glass plates for simulating telescope aberrations, and
(d) a high-gain image intensifier and a digital CCD camera for data recording. Neutral density filters were used to reduce the light intensity since we wanted to simulate severe photon noise.

   Fig. 3 shows a deconvolution experiment with data recorded in the optical setup. The figures show (a) the object (a star cluster), (b) the image degraded by the simulated aberrations and photon noise (about 6000 photon events per image or only 60 photon events in the brightest pixel), (c) the high-resolution image reconstructed from Fig. 3b by inverse filtering, (d) the image reconstructed by the van-Cittert method, and (e) the image reconstructed by the Gerchberg method. Fig. 4 shows a similar experiment with a galaxy as object and photon noise of about 12 000 counts per image.

4. Roll Deconvolution

Roll deconvolution is a method which uses two HST images taken at two different roll angles of the HST (Lohmann and Weigelt 1979, Müller and Weigelt 1987). The roll technique has the advantage that it rotates the psf and also the optical transfer function (Fourier transform of the psf) relative to the astronomical object. A rotation of the transfer function is very useful since it shifts the locations of the zeros in the transfer function. In this way the well-known zero problem of inverse filtering can be overcome completely. This is very important at shortest UV wavelengths where the psf core breaks into speckles and, therefore, there are zeros in the optical transfer function. We have performed computer simulations of the roll deconvolution method with various objects, psfs, and with various degrees of photon noise (50, 100, and 500 photon events per pixel) (Müller and Weigelt 1987).

References

Gerchberg, R.: 1974, *Opt. Act.*, **21**, 709

Lohmann, A.W., Weigelt, G.: 1979, "Image Restoration of ST Photographs", in: Proc. of the ESA/ESO workshop on Astronomical Uses of the Space Telescope, eds. F. Macchetto, F. Pacini, M. Tarenghi, Geneva, 12-14 Feb. 1979, p. 353

Müller, M., Weigelt, G.: 1987, *Astr. Astrophys.*, **175**, 312
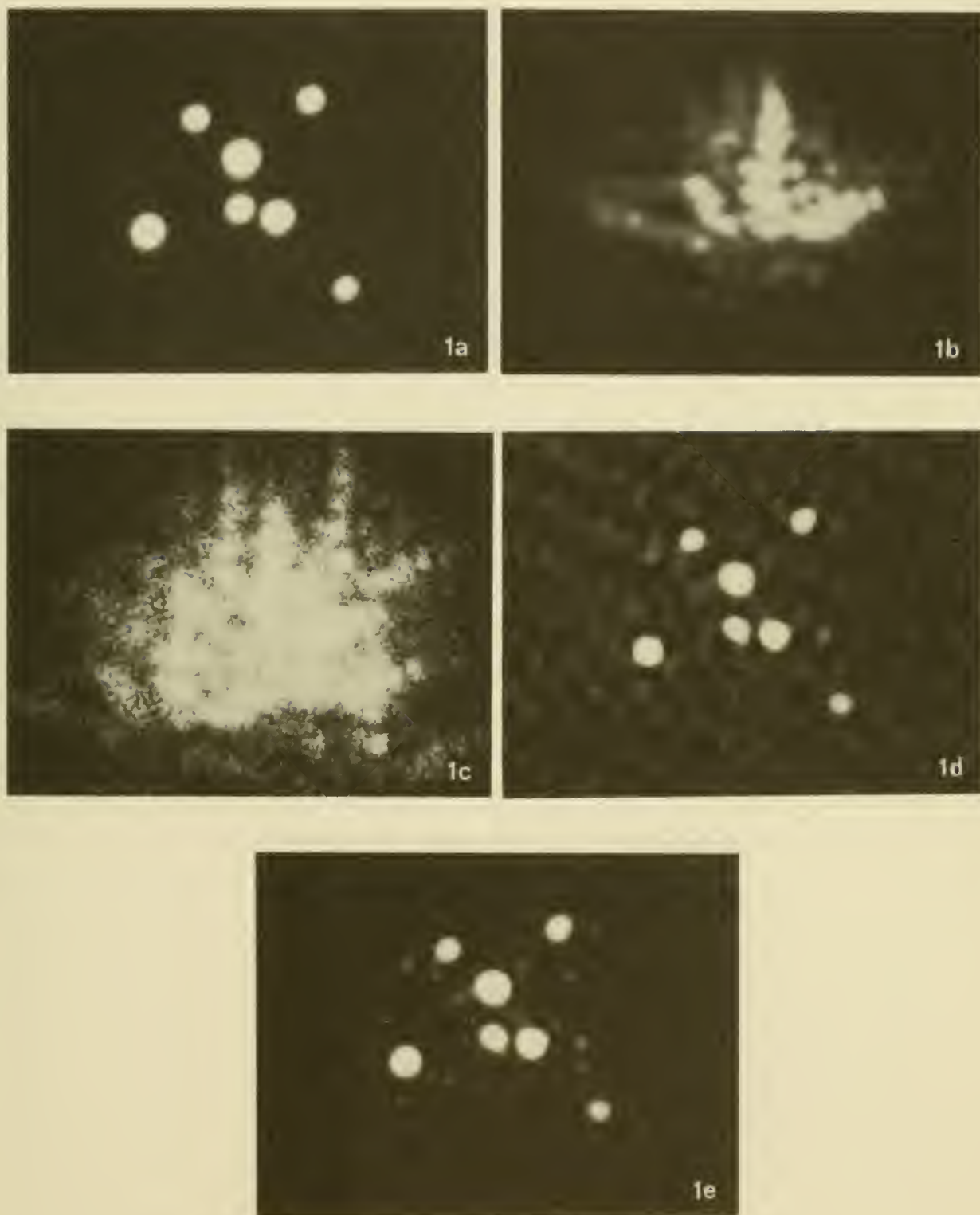
van-Cittert, P.H.: 1931, *Z. Phys.*, **69**, 298

Fig. 1. Computer simulation of the deconvolution of simulated HST images: (a) object; (b) psf; (c) image of Fig. 1a degraded by simulated optical aberrations and photon noise (50 000 photon events per image); (d) high-resolution image reconstructed from Fig. 1c by inverse filtering; (e) image reconstructed by the Gerchberg method.
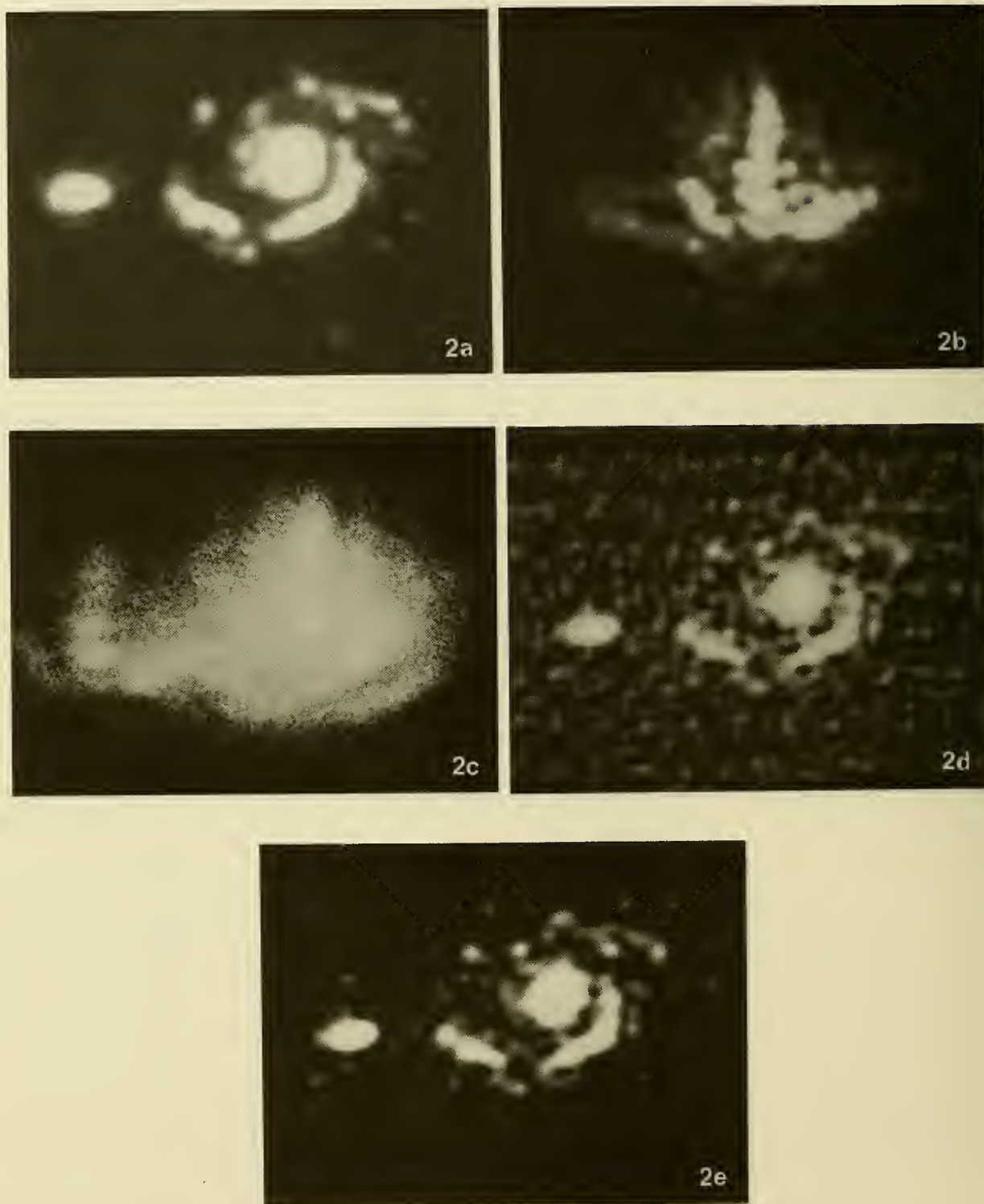
Fig. 2. Computer simulation of the deconvolution of simulated HST images: (a) object; (b) psf; (c) image of Fig. 2a degraded by simulated optical aberrations and photon noise (200 000 photon events per image); (d) high-resolution image reconstructed from Fig. 2c by inverse filtering; (e) image reconstructed by the Gerchberg method.
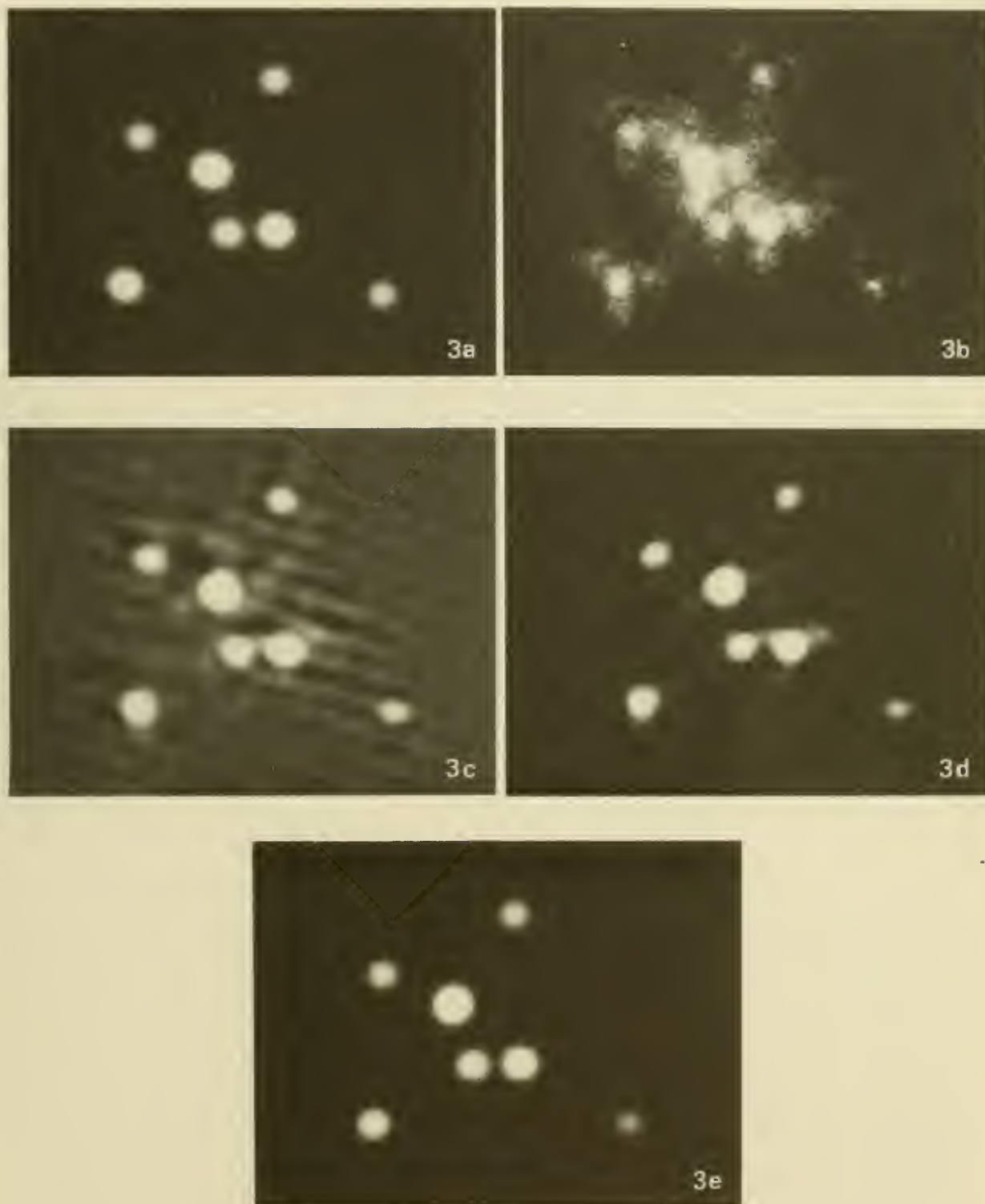
Fig. 3. Laboratory simulation of the deconvolution of simulated HST images: (a) object; (b) image of Fig. 3a degraded by simulated optical aberrations and photon noise (6000 photon events per image); (c) high-resolution image reconstructed from Fig. 3b by inverse filtering; (d) image reconstructed by the van-Cittert method; (e) image reconstructed by the Gerchberg method.
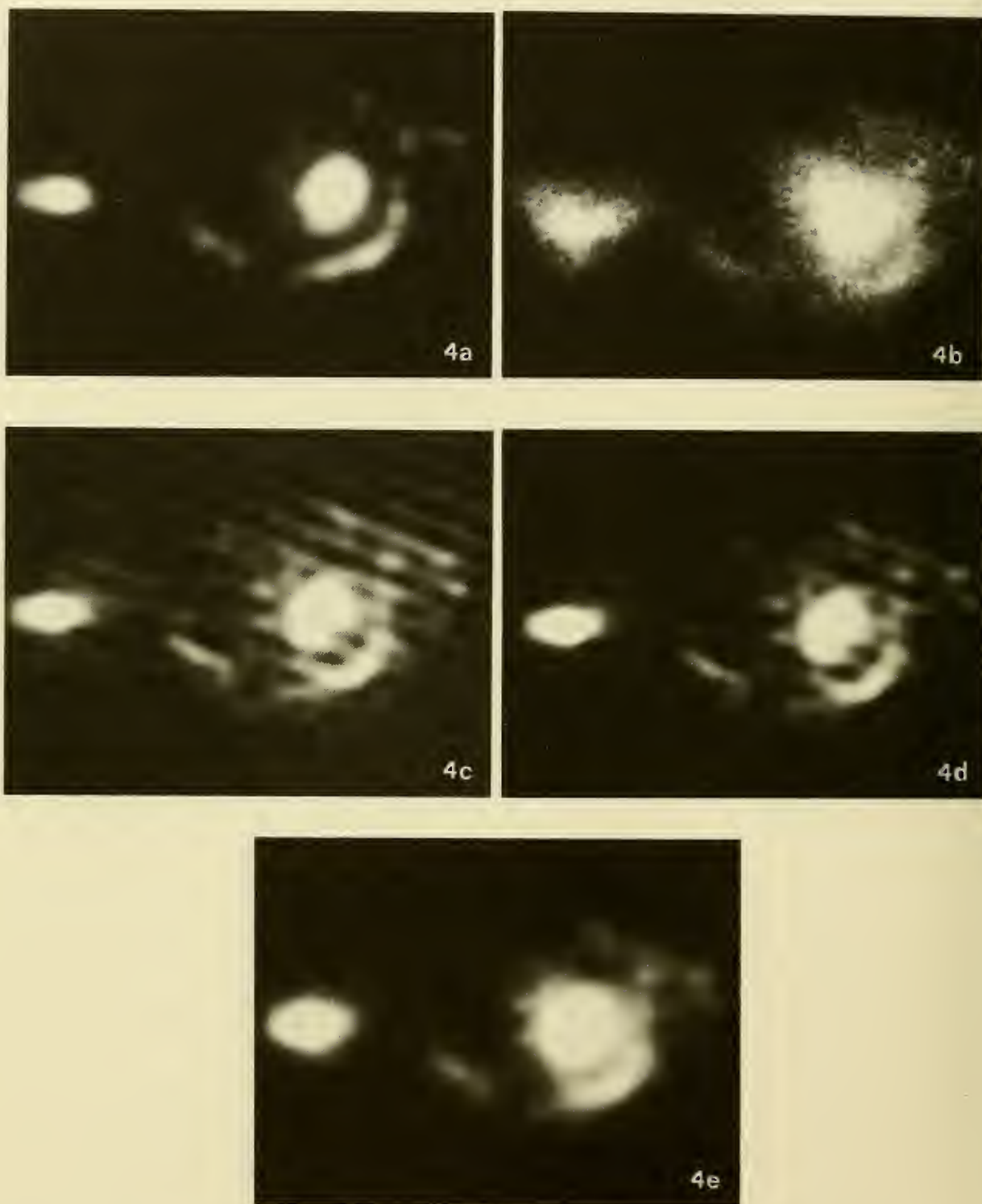
Fig. 4. Laboratory simulation of deconvolution of simulated HST images: (a) object; (b) image of Fig. 4a degraded by simulated optical aberrations and photon noise (12 000 photon events per image); (c) high-resolution image reconstructed from Fig. 4b by inverse filtering; (d) image reconstructed by the van-Cittert method; (e) image reconstructed by the Gerchberg method.

Image Construction from the IRAS Survey and Data Fusion*

Tj.R. Bontekoe
Wyoming Infrared Observatory
University of Wyoming
Laramie WY 82071, USA.

August 21, 1990

Abstract

The IRAS survey data can be used successfully to produce images of extended objects. The major difficulties, viz. non-uniform sampling, different response functions for each detector, and varying signal-to-noise levels for each detector for each scan, have been resolved. The results of three different image construction techniques are compared: co-addition, constrained least-squares, and maximum entropy. The maximum entropy result is superior. We present an image of the galaxy M51 with an average spatial resolution of 45 arc seconds, using 60 micron survey data. This exceeds the telescope diffraction limit of 1 minute of arc, at this wavelength. Data fusion is a proposed method for combining data from different instruments, with different spatial resolutions, at different wavelengths. Direct estimates of the physical parameters, temperature, density and composition, can be made from the data without prior image (re-)construction. An increase in the accuracy of these parameters is expected as the result of this more systematic approach.

1 Introduction

The Infrared Astronomical Satellite (IRAS) surveyed about 95% of the sky, in four broad spectral bands centred on 12, 25, 60, and 100 microns, during a ten month period in 1983. Precession at a rate of about 1° per day, kept the orbit of the spacecraft remaining perpendicular to the earth-sun vector (Figure 1). A semi-overlapping scan strategy was used for the 'all-sky' survey. Redundant coverage on the time scale of hours was provided by advancing the instrument in elongation by half of the width of the focal plane on a subsequent scan, usually the next orbit. For each spectral band there were two detector arrays (Figure 2). The arrays were arranged such that the second one scanned the same area of sky some 5 to 10 seconds later than the first. Different scans over the same area usually intersect at an angles due to precession of the satellite orbit. In addition, most of the scans were taken along small circles. Therefore, even small areas of the sky can be very unevenly covered.

---

*Talk presented at the Workshop on Restoration of HST Images and Spectra, August 21–22 1990, Baltimore.

The nature of the IRAS data is a collection of detector scans cross-cutting the sky at various angles. Consequently, the data are not in the form of an image and require non-traditional methods of image reconstruction. The term 'image reconstruction' actually does not apply; the procedure should be called *image construction*, because the true scene is unknown. Traditional reconstruction routines start with images on an evenly spaced rectilinear grid, of objects convolved with a single point-spread function.

Each of the 59 active IRAS detectors have different Response Functions (RFs). Most detectors have rectangular apertures of 0.75–3 (in the in-scan direction) to 4.5–5 arc minutes (in the cross-scan direction). Consequently, the spatial resolution is different in the two directions. It is possible to improve on the spatial resolution in the cross-scan direction because of the confirmation strategy, and because the two rows of detectors for each wavelength band are shifted by half a detector length in the cross-scan direction.

The detector outputs of two consecutive scans over the galaxy M51 at 60 microns wavelength are shown in Figure 3. This figure can be regarded as ruled-surface plots of the area. The signal is shifted between the two plots due to the half focal plane offset between scans. The main galaxy and its satellite NGC 5195 are clearly resolved. Figure 4 shows the centres of the individual sample positions of the M51 area, together with the outline of a standard 60 microns detector.

2 Data representation

Suppose datum d_n is the calibrated value of the n-th sample, taken by detector number i with its centre at the position (x_n, y_n). Detector i has response function $R_i(x, y)$. Note that in general the RF profile is rotated according to the scan angle, as in Figure 4. Assume that for each sample this rotation has been taken into account in a temporary re-definition of $R_i(x, y)$, now with axes parallel to the axes of the desired map. The measured datum is now the result of a two dimensional integral of the sky brightness $b(x, y)$ and $R_i(x, y)$:

$$d_n = \int dx dy\, R_i(x_n - x, y_n - y)\, b(x, y) + n_n,$$

(1)

with n_n the noise in this n-th datum. Strictly, this is not a convolution since d_n is a single number, and not a function in the continuous variables x and y; therefore it is called a sample.

Digitizing the brightness distribution in pixels reduces the integral to a summation:

$$d_n = \sum_{m=1}^{M} r_{nm}\, b_m + n_n,$$

(2)

with

$$r_{nm} = \int_{\text{area pixel } m} dx dy\, R_i(x_n - x, y_n - y).$$

(3)

Equation 2 can be interpreted as one equation in M unknown discrete brightnesses b_m, with known coefficients r_{nm}. There are in total N samples falling (partly) inside the map, forming a set of N equations in M unknowns (e.g. in Figure 4 $N \simeq 700$). This can be written as the matrix equation

$$d = R_{NM}\, b + n,$$

(4)

96

with $d = (d_1, \cdots, d_N)$ and $n = (n_1, \cdots, n_N)$, the data and the noise vector, respectively. The matrix R_{NM} is the $N \times M$ response matrix, in which the nth row is the set of coefficients R_{nm} for datum d_n. The unknown sky is represented by $b = (b_1, \cdots, b_M)$, a vector of length M, obtained through stacking of the rows (or columns) of the desired map. The RFs are normalized to unit volume.

For samples which overlap the map boundary, it is assumed that the mean intensity just outside the map is the same as just inside the boundary. In this way boundary effects can be apodized.

Formally, the standard deviation n_n of each datum d_n is separately represented in Equation 4. The noise level is estimated from the available data, viz. signal plus noise, using a zero-sum filter. Differences in noise level of a factor three for the same detector, in consecutive orbits have been measured.

Summarizing, the non-uniform sampling, the different RF for each detector and their arbitrary rotations are incorporated in the response matrix. The different noise levels per detector and per scan are represented in the noise vector. The solution b is defined on a regular pixel grid. Since the individual response functions are used, no special provisions are necessary for the small detectors. The image construction has become a numerical mathematical problem which can be solved by many different methods.

Three methods have been compared, co-addition, constrained least-squares, and maximum entropy (see Bontekoe et al., 1991). The best results are obtained with maximum entropy, and are presented here.

3 Image construction

Even when the response matrix R_{NM} is perfectly known, the recovery of the original scene is mathematically impossible. Given the data d, the solution b is not unique. The problem is called *ill-posed*. In addition, the image (re-)construction problem is also *ill-conditioned*. The 'signal-to-noise' ratio of the final map is usually orders of magnitude worse than of the input data. Nevertheless, astronomically meaningful images can be obtained through regularization of the problem; maximum entropy is such a regularization.

In addition to solving for $\hat{b}$ the standard deviation $\sigma_{\hat{b}}$ is also computed for each pixel, and serves as an error map. Although this ignores the covariance of neighbouring pixels, we use this error map to give us some indication of the reliability of features in the solution $\hat{b}$.

The images are constructed on a grid of 60×60 pixels, for the same map area as Figure 4 (15×15 arc minutes). Since there are $N \simeq 700$ samples and $M = 3600$ unknowns, the system of equations is underdetermined.

The MEMSYS3 software package of Skilling (1989) and Gull (1989) is used to obtain a maximum entropy estimate $\hat{b}$ from M51 data. Until recently, the common procedure was to maximize the entropy under the constraint that the goodness-of-fit statistic equals the number of data points. This approach, however, has two major drawbacks. First, within this frame-work there is no consistent method to estimate the error map $\sigma_{\hat{b}}$, although the maximum entropy solution $\hat{b}$ can be found straightforwardly. Second, the $\chi^2 = N$ criterion gives no allowance for the fact that significant data will generate structure in the image $\hat{b}$. This structure is like a set of parameters being fitted from the data. The effective number of parameters G supporting the underlying the structure should be

subtracted from N to obtain the number of degrees of freedom with which to compare χ^2 (see also the discussion in Gull, 1989).

The Bayesian estimate for the present case is $G \simeq 400$. The remaining $N - G \simeq 300$ is the number of degrees of freedom for this problem, and this should be used for the χ^2 statistic.

The maximum entropy solution shows fine details, especially in areas of low brightness where the other methods fail. Although such maps can be used sensibly, inferences ought to be made by overlaying them with masks and computing integrals over the masked solution.

The Figures 5 and 6 are the result of coarsening the maximum entropy solution, originally computed on 60×60 pixels, to an effective 30×30 grid by using 900 2×2 pixels boxcar masks, being 1 in the square of interest and 0 outside. The 60×60 grid is retained, however, and each group of four pixels is assigned the same value $\hat{b}$. Finally, this coarsened map is smoothed again using a 2×2 boxcar filter.

Division of brightness map by the error map yields Figure 6.

Most of the spiral structure in Figure 5 coincides well with the spiral arms of the Hα image at 8 arc second resolution in van der Hulst et al. (1988). Extensions of the spiral arms at low brightness levels in the maximum entropy solution line up very well with outer parts of the spiral arms in Hα. The brightest point does not coincide with the nucleus in the Hα, but lies about 25 arc seconds to the South. The second source in the centre also has no visual counterpart and coincides with an inter-arm region. The companion galaxy seems resolved into a strong point source, towards its South-East, which coincides exactly with a sharp maximum in Hα and extended structure towards the North and West.

Figure 6 summarizes the maximum entropy result in a statistical sense. The peaks in the galaxy and companion are 8σ and 9σ detections, respectively. The spiral arms are 1–3 σ detections, and the secondary peak in the nucleus of M51 a 3σ detection. A unidentified source near the Northern boundary of the map is a 4σ detection, although its brightness is low. This might be an artifact from the treatment of the boundary.

A difficult issue is the final spatial resolution in the map. Since MEMSYS3 finds $G \simeq 400$, it is tempting to distribute this number evenly over the image as the number of independent picture elements, yielding an average spatial resolution of 45 seconds of arc. This is an improvement above the diffraction limit of the telescope as well as the classical limit imposed by sampling theory, both of which are 1 arc minute. Spatial resolution, however, is dependent on the 'signal-to-noise' in the original data and consequently non-uniform over the map. Therefore the 400 fitted parameters can not be evenly distributed, and areas of high brightness can have a better spatial resolution than average. The reverse is true for low brightness areas.

Although the IRAS survey mission was not intended to produce images, the major difficulties, viz. non-uniform sampling, different RFs for each detector, and different signal-to-noise levels for each detector for each scan, have been dealt with. Astronomically meaningful images can be produced of the far-infrared sky from the IRAS survey data, but they have to be constructed by advanced numerical techniques. Overall, the MEMSYS3 result is superior to the others. The images produced by MEMSYS3 show better spatial resolution, are non-negative and show plausible structure even at low brightness levels.

4   Data Fusion

One of the central questions in astrophysics is the correct identification of radiative processes operating in a given source. A fresh attack is proposed on the classification of physical processes in a complex field from a fundamentally new direction. The goal is to take data of a given field, obtained at a variety of wavelengths and spatial resolutions, and produce *images* of operative *physical* processes and the corresponding parameters.

First the data acquisition process is briefly described. Emphasis is given to the correct dimensions of all quantities. The sky brightness $B(\theta, \phi, \lambda)$ (in $Wm^{-2}m^{-1}sr^{-1}$) is a function of position of the sky (θ, ϕ) and of wavelength λ. When the instrument is pointed in the direction (θ_i, ϕ_i), the response function, representing the blur, is defined $R(\theta, \phi, \theta_i, \phi_i)$ (dimensionless). In addition the blurred signal must pass a colour filter $F(\lambda)$ (dimensionless). Each datum of measured flux d_i (in Wm^{-2}) is now the result of both 'convolutions',

$$d_i = \int\int d\theta d\phi \, R(\theta, \phi, \theta_i, \phi_i) \int d\lambda \, F(\lambda) \, B(\theta, \phi, \lambda). \tag{5}$$

The important fact to notice is that the data acquisition is assumed to be linear with the input signal, which is the essential task for the calibration. It is this linearity that allows decomposition of the sky brightness into various components. Note also the difference between B here and b of the previous sections; the latter is the brightness integrated over the colour filter. The effect of noise is well understood in linear problems, and does not affect the theoretical analysis.

The ensemble of number densities $n_s(r, \theta, \phi, T_s)$ (in m^{-3}) of sources, indexed with s, can be estimated, when the source's radii R_s (in m), and template spectra $I_s(\lambda, T_s)$ (in $Wm^{-2}m^{-1}sr^{-1}$) are given. The best fitting temperature T_s (in degrees K) of the various n_s results as part of the solution. An infinitesimal volume $dv = r^2 dr d\Omega$ (in m^3), radiates with a spectral power per unit wavelength dw (in Wm^{-1}) equal to

$$dw(r, \theta, \phi) = dv \sum_s n_s(r, \theta, \phi, T_s) \, 4\pi R_s^2 \, I_s(\lambda, T_s), \tag{6}$$

assuming isotropic radiation by all sources. The contribution to the sky brightness of dv is

$$dB(\theta, \phi, \lambda) = \frac{dw(r, \theta, \phi)}{4\pi r^2 d\Omega} e^{-\tau(r, \lambda)}, \tag{7}$$

where $\tau(r, \lambda)$ is the optical depth (dimensionless). Integrating over the line of sight, the sky brightness becomes

$$B(\theta, \phi, \lambda) = \int_0^\infty dr \, e^{-\tau(r, \theta, \phi, \lambda)} \sum_s n_s(r, \theta, \phi, T_s) \, R_s^2 \, I_s(\lambda, T_s). \tag{8}$$

Under the condition that the absorption, represented by τ, and the emission I_s are independent of the local radiation field, the sky brightness B is a linear function of the densities n_s.

Since the data acquisition is assumed linear with brightness, there is linear relation between the data d_i and n_s the physical parameters of the distribution of matter in space. It requires simultaneous solution of the equation of radiative transfer and the instrumental inversion. However complicated, the problem is linear, and linear problems can be solved by many numerical techniques.

The critical factor in resolving three-dimensional structures in the universe is knowledge of the factor $\tau(r,\theta,\phi,\lambda)$. The full solution $n_s(r,\theta,\phi,T_s)$ for all sources describes the composition, temperature, and density of all visible matter. If this information is not present in the data, upper bounds for the desired densities are still a useful result.

If one is less confident in the quality of the data or ones knowledge of the absorption, one can apply a two dimensional version of the theory in which only projected densities are defined.

Acknowledgement

References

Bontekoe, Tj.R., Kester, D.J.M., Price, S.D., de Jonge, A.R.W., Wesselius, P.R., 1991, to be submitted to Astron. Astrophys.

Gull, S.F., 1989, in Maximum Entropy and Bayesian Methods, Cambridge 1988, ed. J. Skilling, p. 53, (Kluwer).

Skilling, J., 1989, in Maximum Entropy and Bayesian Methods, Cambridge 1988, ed. J. Skilling, p. 45, (Kluwer).

van der Hulst, J.M., Kennicutt, R.C., Crane, P.C., Rots, A.H. 1988, Astron. Astrophys. **195**, 38.

Figure captions

Figure 1: Schematic drawing of IRAS orbital geometry.

Figure 2: IRAS focal plane. Of the 62 infrared detectors the 3 filled-in were inoperative; the cross-hatched detectors showed a higher noise level. The normal scan direction of images is shown.

Figure 3: Detector scans from two consecutive scans covering the galaxy M51 (60 microns). The scan direction is from right to left and each has a length of 0.5 degree. The detector scans are displaced vertically corresponding to their cross-scan position in the focal plane. The figures can be regarded as ruled surface plots. The maximum signal in both plots is 430 detector units.

Figure 4: Positions of the samples in the area of M51 (60 microns). The circles represent the two smallest detectors in the band, viz. detectors 11 and 31 (see Fig. 1). The scans roughly run from top-left to bottom-right. The area is covered by about 700 samples. The fat contours outline the RF profile of a normal size detector, at the 90, 50, 10, 2, and 1% level of its maximum. The detector is centred at the position indicated with an asterisk. A grid with a one arc minute spacing is superimposed.

Figure 5: Map of M51 (60 microns) from the MEMSYS3 maximum entropy method. Area is the same as in Figure 4. The lowest contour is at a level 150; subsequent contours with increments of a factor 2.

Figure 6: Signal-to-noise map of the MEMSYS3 solution of M51. Lowest contour is at 1σ, subsequent contours are separated by 1σ.
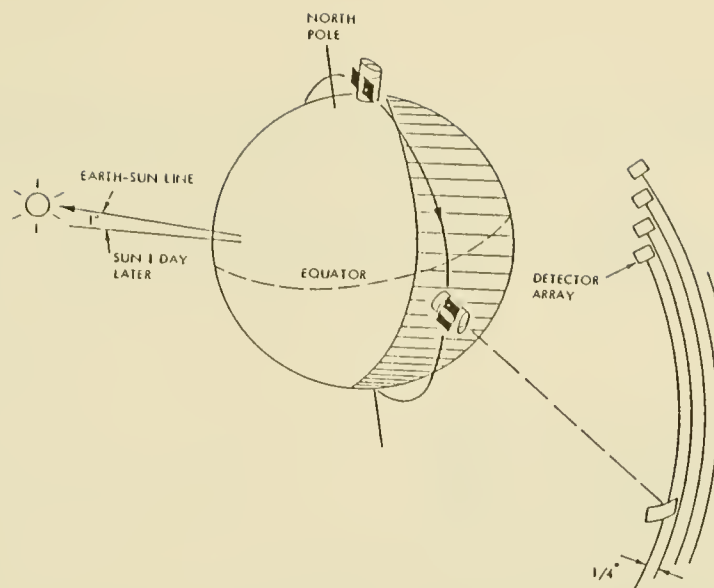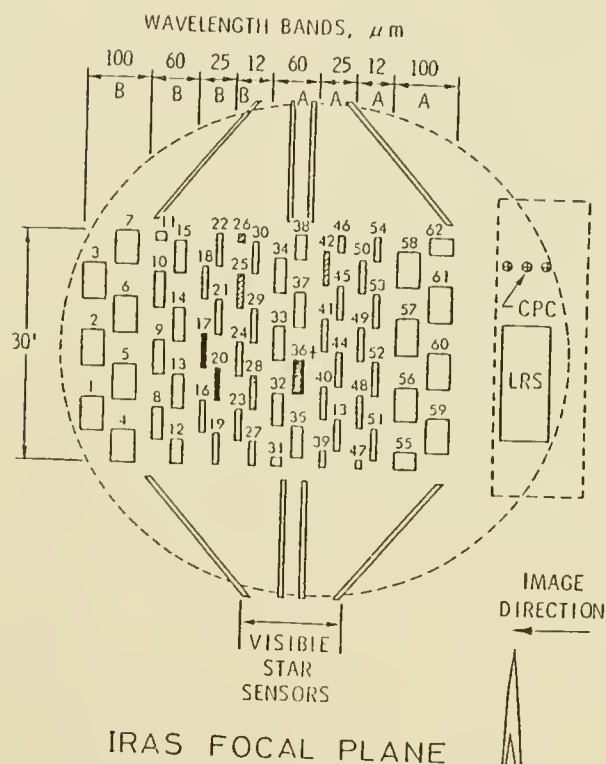
Fig. 1

WAVELENGTH BANDS, μm
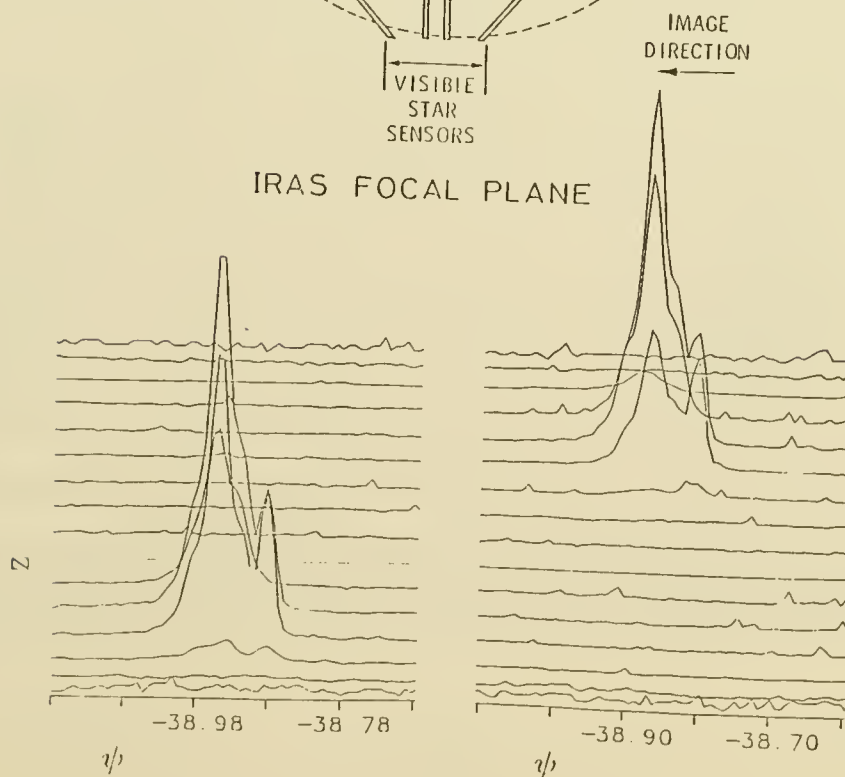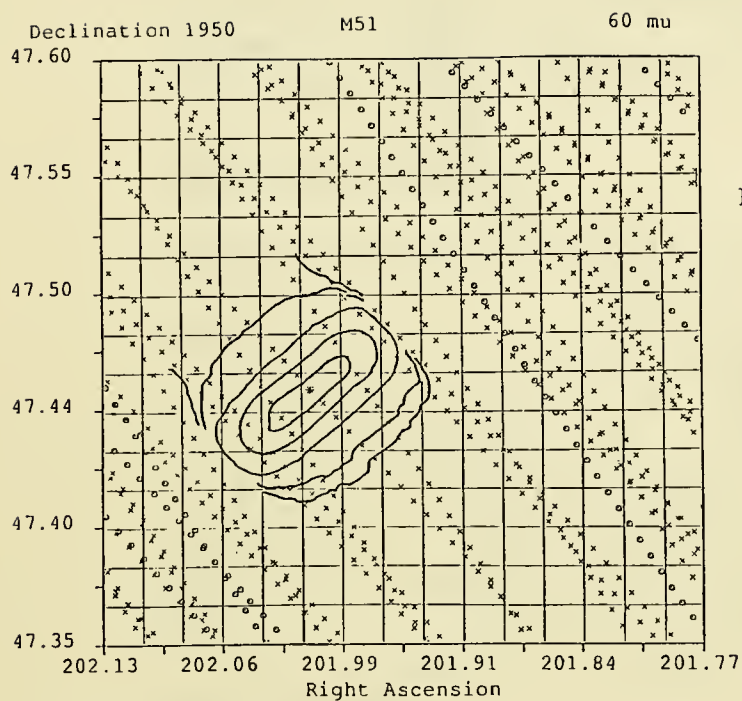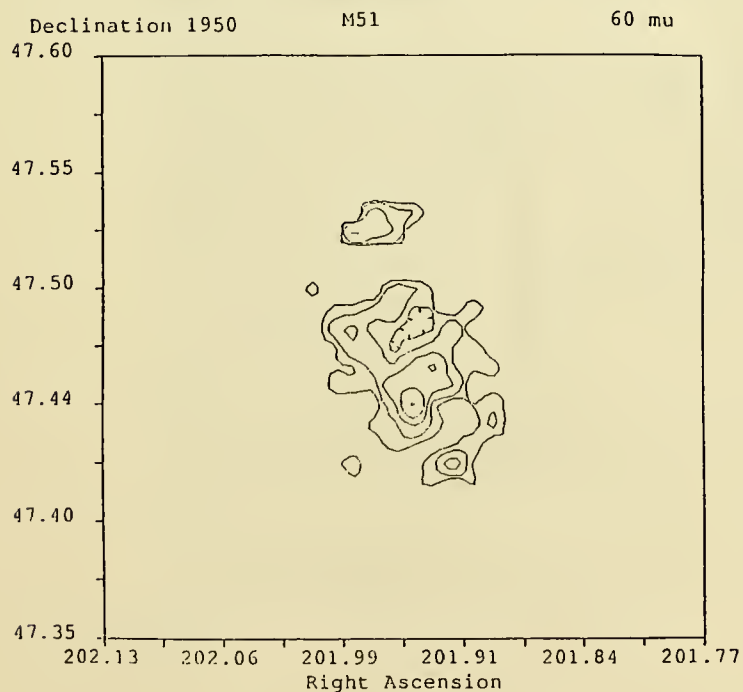
IRAS FOCAL PLANE

Fig. 2
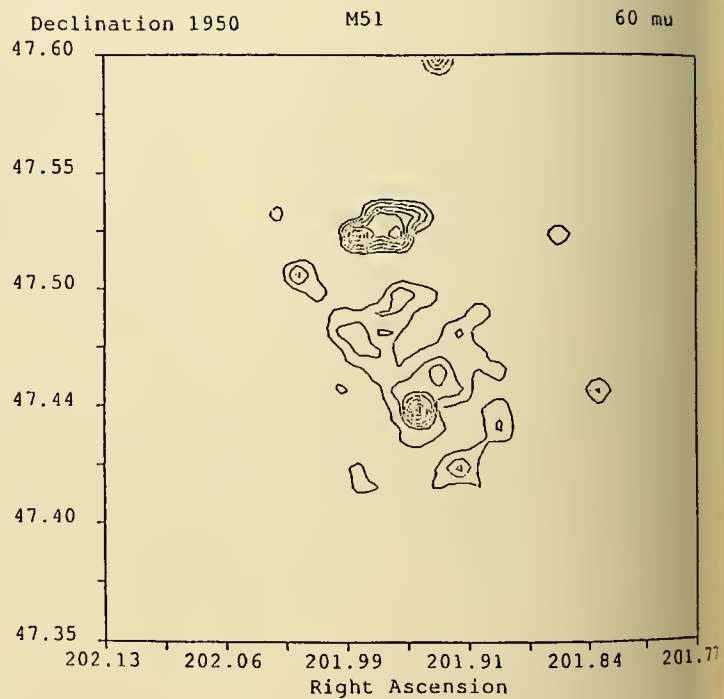
Fig. 3

Fig. 4

Fig. 5

Fig. 6

A Wiener Filter Version of Blind Iterative Deconvolution

P. Nisenson, C.Standley, and D. Gay

Harvard-Smithsonian Center for Astrophysics

I. Introduction

Over the last decade, great progress has been made in high angular resolution imaging at optical and infrared wavelengths. Techniques have been developed which allow nearly diffraction-limited image recovery from images which have been severely degraded by atmospheric turbulence and telescope aberrations. Imaging at radio frequencies has also led to the development of a number of new and powerful image processing algorithms for handling the data from large telescope arrays. Techniques such as CLEAN (Hogbom, 1974), MEM (Gull and Daniell, 1978) and Gerchberg-Saxton (Gerchberg and Saxton, 1972) have proven to be useful, not only for radio map reconstruction, but also for a wide range of other astronomical imaging applications. Blind Iterative Deconvolution (BID) is a technique which was originally proposed for the correction of the effects of atmospheric turbulence on astronomical images. In this technique, both the restored image and the degrading point spread function (PSF) may be recovered from a single high-signal-to-noise ratio short exposure (speckle) image.

At the Center for Astrophysics, we have implemented a modified version of BID using the general approach proposed by Ayers and Dainty (1988). The basic approach is to alternately deconvolve the original data by the PSF and then by the restored image. A set of physical constraints are applied after each iteration. The iterations are continued until an image and PSF are found which give the original data when convolved (with each other) and which adhere to the physical constraints. BID is most useful for cases where

the PSF is poorly known or time dependant. It has a number of features which should be important for a wide variety of scientific problems, such as the blurred images from HST. We have performed a number of numerical and experimental tests with the algorithm and have found that, in many cases, it provides remarkable reconstructions from severely degraded imagery.

II. Blind Iterative Deconvolution

Blind Iterative deconvolution (BID) combines constrained iterative techniques such as those developed for phase retrieval (Gerchberg and Saxton 1972, Fienup 1978) with blind deconvolution (Lane and Bates 1987). One starts with an image which is degraded by some blurring function. A necessary condition for the algorithm to work is that the blurring function be invariant over the entire image field to be restored (stationarity). It is also assumed that the degradation has been a linear operation. The general approach is then to find a pair of functions whose convolution gives the input image within a set of physical constraints. These constraints include positivity in the two convolved functions, the image and psf support (non-zero) region and the signal-to-noise ratio in the Fourier transform (FT). While it has not been proven that the derived functions are unique, complicated images appear to converge on only one sensible solution. A flow diagram for the technique is given in figure 1. One starts with a degraded image and an initial estimate of the point spread function (PSF). The initial PSF can be randomly chosen, however the number of iterations required to converge on an acceptable answer is highly dependant on how close the first estimate of the PSF is to the actual PSF. Both inputs are Fourier transformed and a deconvolution is performed by constructing a Wiener filter from the FT of the PSF. The technique of Wiener (or Optimum) filtering damps the high frequencies and minimizes the mean square error between each estimate and the true spectrum. Denoting the FT by

lower case letters this filtered deconvolution takes the form:

$$o(u, v) = \frac{i(u, v).\phi(u, v)}{p(u, v)}$$

where the Weiner filter used in our computations, $\phi(u, v)$, is given by:

$$\phi(u, v) = \frac{p(u, v).p^*(u, v)}{|p(u, v)|^2 + |n(u, v)|^2}$$

$p(u, v)$ and $n(u, v)$ are the PSF and noise spectra respectively. $n(u, v)$ usually can be replaced with a constant estimated from a high frequency region in the spectrum where the object power is small.

The result is transformed back to image space and positivity and support constraints are applied. After the support constraint is applied, the negatives in the image are set to zero. The negatives are then summed and uniformly subtracted within the support region in order to preserve the total power in the image. After subtraction, some areas of the image may become negative. If this is the case, the negatives are again truncated, summed and subtracted. This procedure is repeated until the restored image is all positive. The ratio of positives to negatives in the image is also used as a diagnostic of convergence. The FT of the original degraded image is then deconvolved by the FT of the restored image obtained from the first iteration. The result is transformed back to image space. Again, positivity and support constraints are enforced. The result is a new estimate of the PSF. The iteration continues until a stable solution is found. A damping factor is used to stabilize the iteration, particularly important when the PSF estimate is still inaccurate. About 20% of the image (or PSF) from the previous cycle is averaged with the new image (or PSF) in the early stages of the process. This percentage is reduced when the iteration has nearly converged.

Two criterion have been found to be very useful in determining the completion of the iteration: the ratio of positive power to negative power in the restored image and psf; and the rms difference between iterations. Both criteria drop irregularly in the first few

cycles of the iteration, but they both level off and stabilize when the operation is close to convergence. After examination of the output image and the PSF, the results may be fed back into the loop for continued iterations.

There are a number of parameters which must be chosen in order to ensure convergence and an optimum result. Probably the most important are estimates of the signal-to-noise ratio in the data for construction of the Wiener filter and the region of the support constraint. It is also very important that the image and PSF remain aligned with the support constraint, to avoid truncation. This is done by centering the initial image and PSF, calculating the two support regions, and then recentering the PSF after each iteration.

III. Reconstructions Using Iterative Deconvolution.

The work on BID already undertaken at CfA has produced an algorithm that has been tested with computer simulations and also applied to some real data. Results of the simulations are shown in Figs. 2. Fig. 2a shows the input diffraction limited image of 8 point sources, the bottom right- hand "point" being two unresolved points. This image was convolved with the PSF in Fig. 2b (a simulated atmospheric PSF) and then degraded by photon noise. The level of photon noise was set by assuming that the image was recorded with a one second exposure with a 2.4 meter telescope, that the stars were 12th magnitude, that the detector had 30% quantum efficiency and that the optical efficiency was 50%. The resulting input (speckle) image is shown in Fig. 2c. The starting guess for the PSF (Fig. 2d) used in the first cycle of BID was a gaussian with random noise and a half power width approximately the same as the "seeing". In most real situations, there is usually some reasonable estimate of the PSF which, when used as a first guess, should improve the rate of convergence. Here, a randomly chosen PSF was used to demonstrate the dramatic evolution of the reconstructed PSF towards the actual degrading PSF, despite the quasi-random starting estimate. The image and PSF obtained from BID after 250 cycles are shown in Figs. 2e and 2f. Comparison of these with Figs. 2a. and 2b. show a dramatic

recovery of both the morphology and relative intensities present in the diffraction limited image and PSF. The dramatic convergence towards a solution is seen in Fig 3. Here the percentage of negatives is plotted against the number of iterations. After 180 iterations BID has converged on a stable solution. However, a roughly constant percentage of negatives remains, which adds a constant level to the reconstructed image. Rescaling the restored image to preserve the total integrated power in the original image removes this level.

Some initial attempts have been made to use the technique on real optical CCD data. Figs. 4a and 4b show results of BID processing on CCD images of supernova SN1987A recorded with a 30Å wide, Hα filter at the CTIO 4-meter telescope in January, 1990. Fig. 4a shows the original image of the supernova, its two companions, and the extended nebulosity during the pre-supernova red supergiant phase of the precursor star. The wide companion is separated by 2.9 arcseconds from the SN. Fig. 4b is the BID reconstruction, using a single star in a close field as the first guess at the PSF, after only 10 cycles of processing. Clearly the image has been sharpened, though only limited resolution recovery is possible due to the long exposure atmospheric transfer function cutoff. These results demonstrate the power of the technique, and Fig. 4c shows that one obtains rapid convergence if a good initial estimate of the PSF is available. Excellent results have also been obtained on X- ray (Standley et al, 1990) and EUV Solar images from Skylab (Karovska and Habbal, 1990). Tests of BID on WF/PC simulation data were also impressive in their recovered resolution and fidelity to the original data.

IV. Summary

Blind Iterative Deconvolution shows great promise as an image enhacement technique for astronomical images with unknown or poorly known degradations. BID provides not only improved spatial resolution but also allows extraction of the PSF of the degrading process. Since it always Wiener filters the original data, it never walks away from a data dependant, linear solution. A detailed comparison of BID with MEM and other iterative

deconvolution techniques is also of great interest. BID may prove to have important application to the processing of HST images (and spectra) since precise measurements of the PSF for all fields may be difficult. While initial tests indicate that it degrades gracefully with noise (producing a noisey, but undistorted image), detailed characterization and rigorous analysis of the technique is required before it can be used for scientific purposes.

References.

Ayers, G.R. and Dainty, J.C. 1988, Opt. Lett., **13**, 547.

Chanan, G.A., Helfand, D.J., and Reynolds, S. 1984, Ap. J. Lett., **287**, L23.

Fienup, J.R. 1978, Opt. Lett., **3**, 27.

Gerchberg,R.W. and Saxton, W.O. 1972, Optik, **35**, 237.

Gull, S.F. and Daniell, G.J. 1978, Nature, **272**, 686.

Hogbom, J.A. 1974, Astr. Astrophys. Suppl., **15**, 417.

Karovska, M. and Habbal, S. 1990, Ap. J. Lett., in press.

Lane, R.G. and Bates, R.H.T. 1987, J. Opt. Soc. Am. A, **4**, 180.

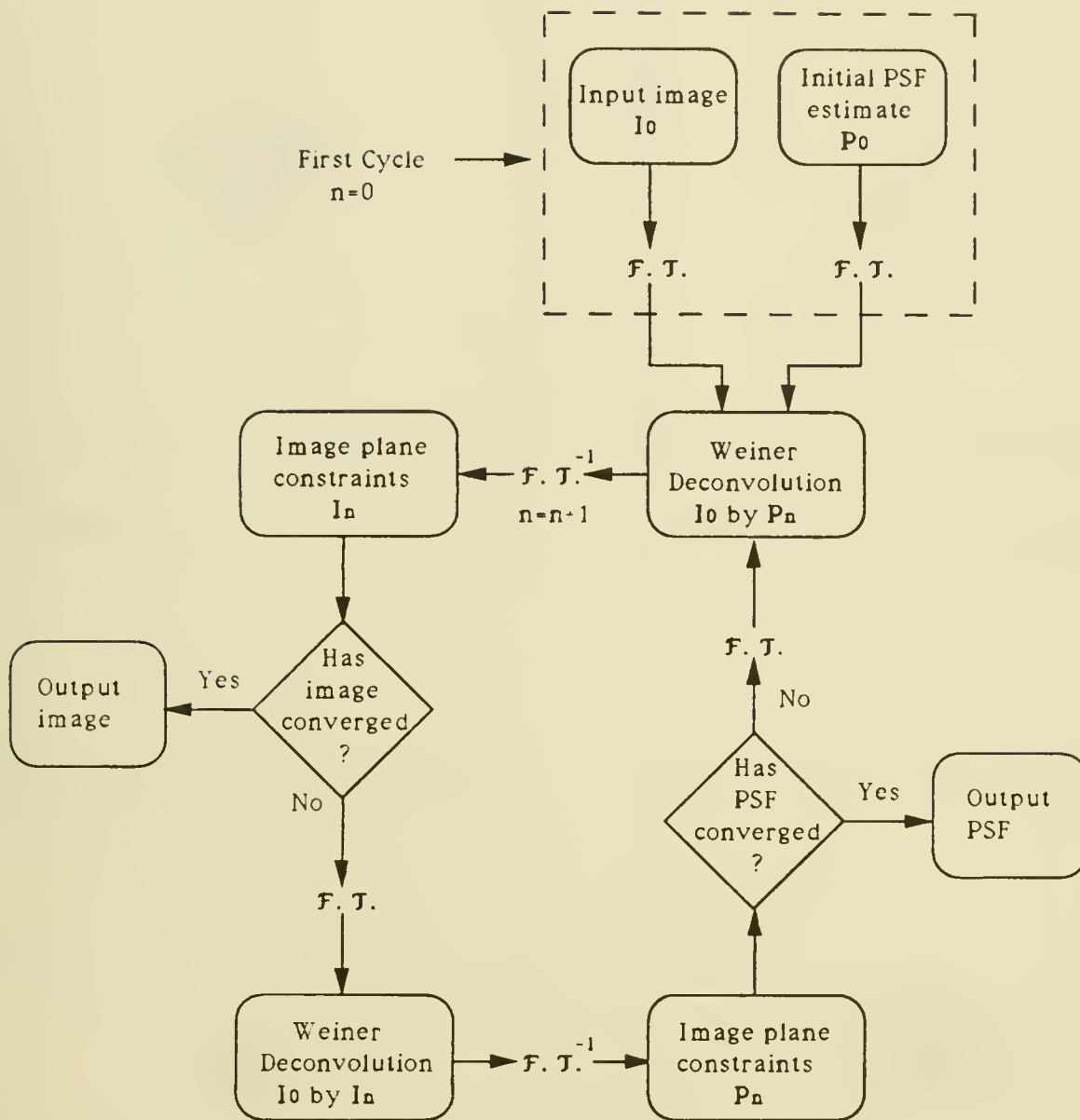Standley, C. and Nisenson, P. 1989, BAAS, **21**, 1072.

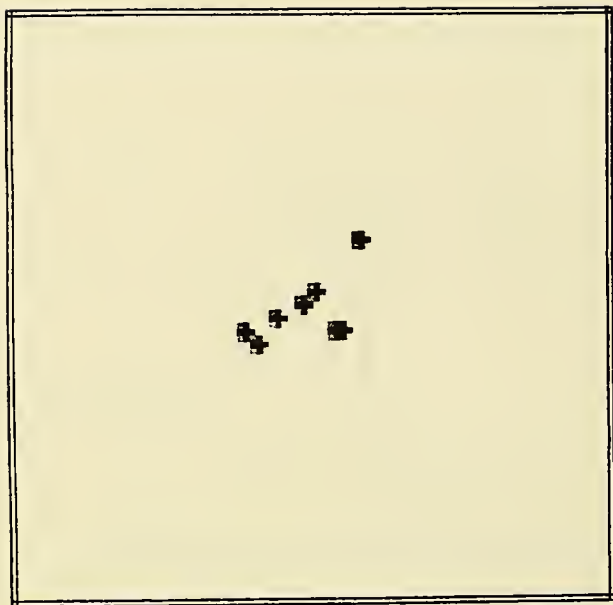Fig. 1 Flow Diagram of the Blind Iterative Deconvolution Algorithm.

Fig. 2a. The Diffraction Limited Input Image.

Fig. 2b. The Degrading PSF

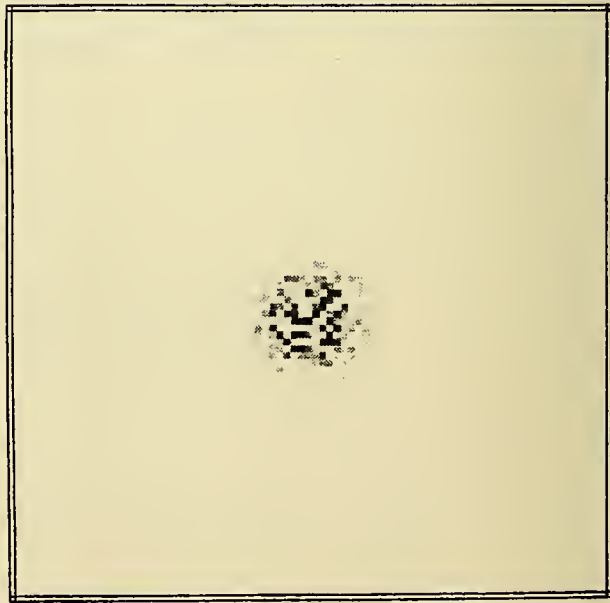Fig. 2c. The Degraded Image. Convolution of 2a. with 2b. Including Photon Noise.

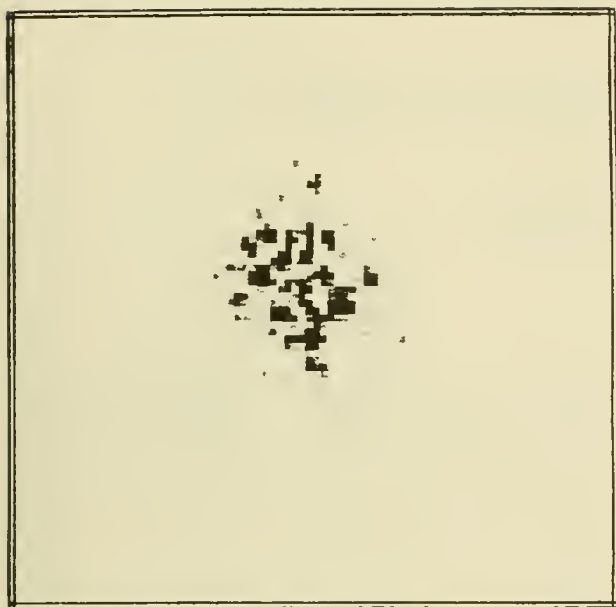Fig. 2d. The Starting PSF (P_0) for BID.

Fig. 2e. The Reconstructed PSF After 250 Cycles. Fig. 2f. The Reconstructed Image After 250 Cycles.
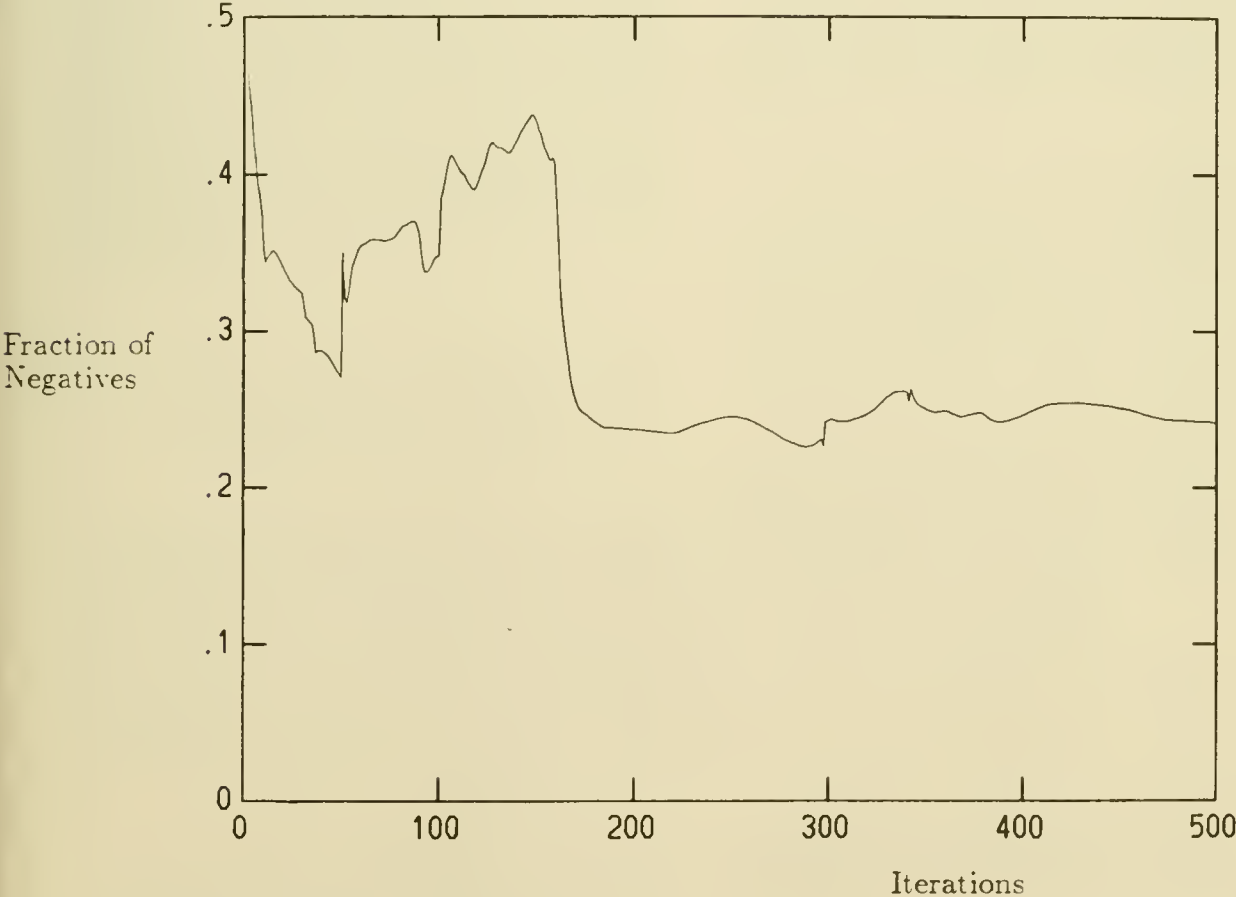

Fig. 3. The Convergence of BID for the 8 Point Image.
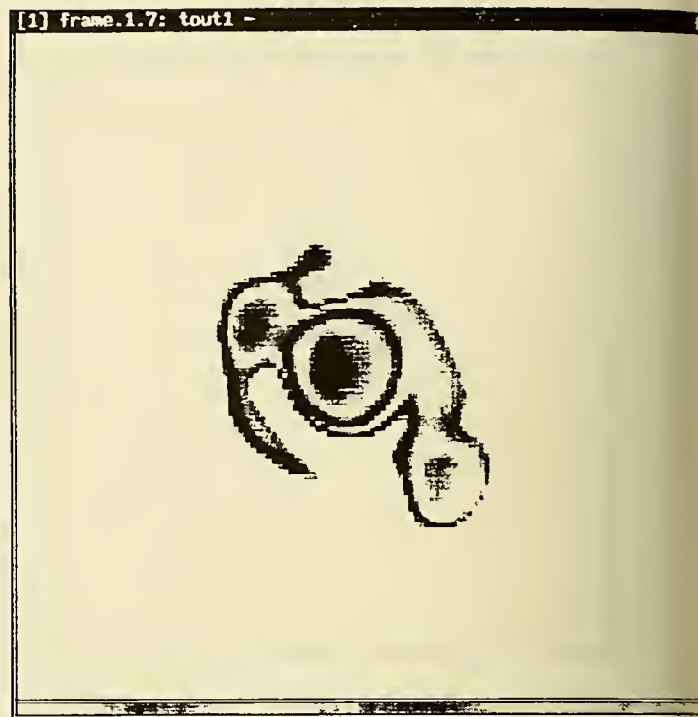
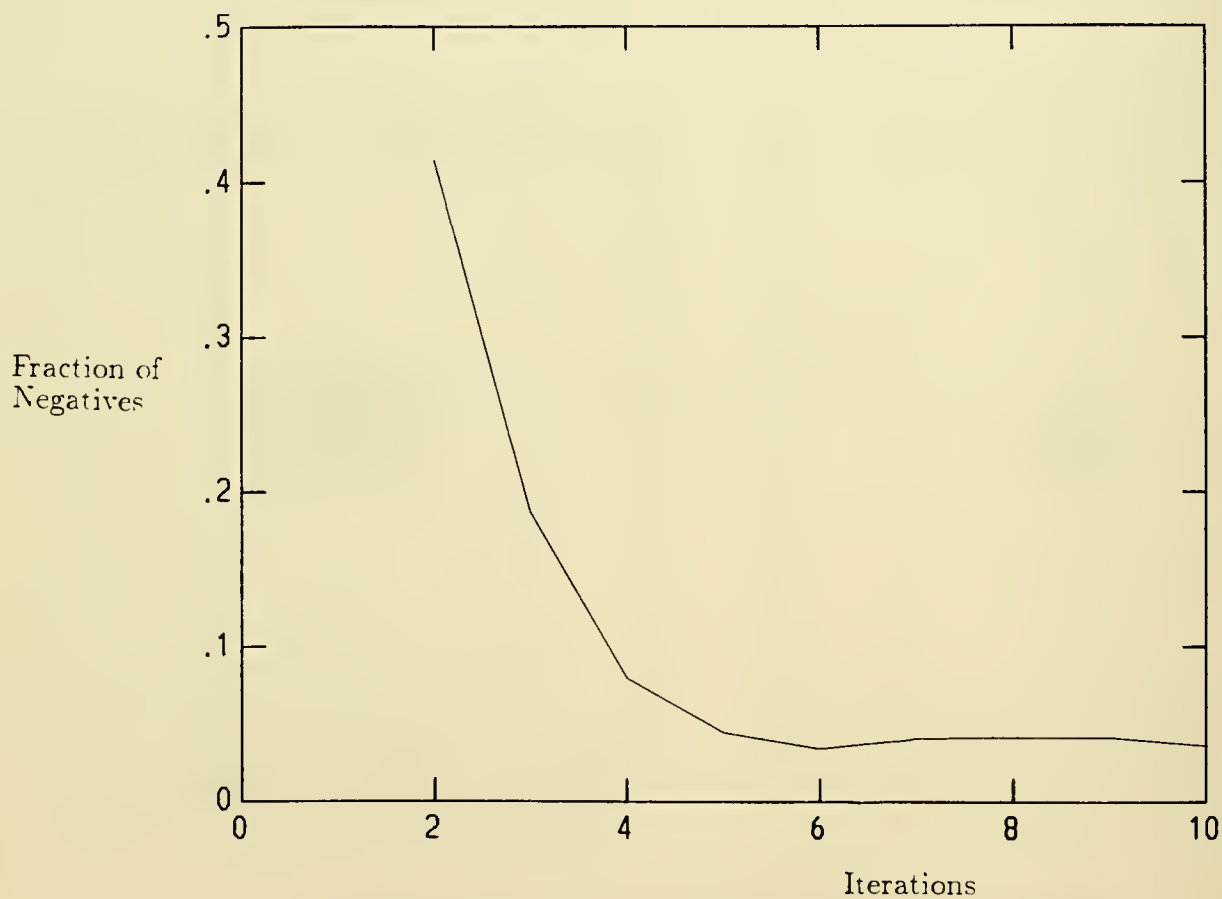Fig. 4a. A 30Å, Hα CCD Image of SN1987A   Fig. 4b. The Reconstruction of the SN1987A Image.

Fig. 5. The Convergence of BID for the SN1987A Image.

Optical Interferometers in Space:

What the Program Can Do for HST;
What HST Can Do for the Program

David L. Meier
Jet Propulsion Laboratory

ABSTRACT

The Hubble Space Telescope is the first imaging optical interferometer to be launched into space. It is of the Fizeau class, forming its fringes in the image plane onto CCD detectors, thereby directly detecting the "dirty image". An unaberrated HST would have had little need for deconvolution of the dirty image with the point spread function. However, with the spherical aberration in the primary, deconvolution and other techniques commonly used in interferometric imaging will be important. Some techniques which have been applied to our imaging simulations of optical interferometers at JPL are described, including image pre-processing which allows the CLEAN algorithm to work well on optical images. Based on some results of this workshop, we suggest a general restoration procedure for images obtained with HST, optical interferometers, and radio interferometers which is a hybrid of CLEAN and the Quantified Maximum Entropy Method. Results of this and similar conferences, and any related documents and software, should be archived and distributed as widely as possible, as they will affect the performance of future NASA missions, such as the imaging optical interferometers.

I. INTRODUCTION TO OBSERVATORY CLASS, SPACE OPTICAL INTERFEROMETERS AS FOLLOW-ON MISSIONS TO THE HST

Interferometers are classified on the basis of their scientific purpose (imaging or astrometric), on the basis of how they perform fringe detection (in the image plane or in the aperture or pupil plane), and on the basis of the observing frequency (optical, radio, etc.), which affects the practical aspects of fringe detection, noise response, etc. As we are primarily concerned with imaging at this workshop, we will not consider astrometric interferometers here in any detail.

A. Imaging Interferometers

Virtually all imaging devices are interferometers. They rely on the wave nature of light to form interference fringes between two or more receiving elements (such as small sections of a parabolic mirror or widely-separated radio antennas). If the waves received by each element are allowed to simultaneously and coherently interfere, with zero relative delay, the resulting fringe pattern is called an image (a "dirty" image to be more precise) and the instrument is called an "all-with-all image plane interferometer", a Fizeau interferometer, or a telescope. If the waves are combined pairwise, rather than simultaneously, the amplitude and phase of each resulting fringe pattern measures one (two-dimensional) Fourier component of the image. The large number of such measurements, distributed over the Fourier plane (often called the "aperture" or "pupil" plane), can be Fourier inverted to yield virtually the same image as detected by the image plane interferometer. The quality of an interferometric image is directly related to the number and distribution of these visibility measurements in the Fourier plane. Filled, or nearly filled, telescopes sample many visibilities and therefore produce quite good dirty images.

Note that some pupil plane optical interferometers employ an all-with-all beam combination scheme. The resulting pattern on the detector is not a direct image of the object, but it can be Fourier inverted, used to extract the fringe amplitudes and match them to each element pair with known phase, rearranged in the pupil plane, and then transformed back to recover the dirty image.

Because of the large wavelength and the necessary detector technology, nearly all interferometers in the radio region form fringes in the aperture plane, whereas both pupil and image plane interferometry is used in optical observations. The advent of focal plane arrays may increase the use of image plane interferometry at far infrared and longer wavelengths.

B. Space-based Interferometry And the Astrotech 21 Program

The goal of NASA's Astrotech 21 program, managed by JPL, is to identify and develop technologies which will be necessary for missions which are to be flown in the next century. Optical interferometry is one of these technologies. Currently Astrotech 21 is looking at three types of optical interferometry missions:

1. single-spacecraft freeflyers,
2. multiple-spacecraft free-flying, long-baseline interferometers,
3. lunar-based, long-baseline interferometers.

All of these involve both imaging and/or astrometric concepts. Single-spacecraft missions are likely to be the first to follow Hubble, so for the remainder of the paper we will concentrate on these only.

Figure 1 shows examples of interferometers studied in Astrotech 21. The GOLAY concept is an example of a dilute aperture or Fizeau class interferometer which forms its fringes in the image plane. Note the secondary mirror and Cassegrain design. The Hubble Space Telescope is a special case of this Fizeau class as it has a nearly-filled aperture. COSMIC is an example of a pupil plane interferometer, similar in style to the many ground-based radio interferometers currently in use. Note the all-with-all beam combination scheme in the pupil plane (see inset in Figure 1b). Figure 1c shows the Precision Optical INTerferometer in Space (POINTS) wide field astrometric concept, and 1d shows the Optical Space Interferometer (OSI) concept which can be used to do both narrow field astrometry and simple imaging by measuring fringe visibilities at a few specific points in the aperture plane.

C. The Fizeau Filled-arm Telescope

One mission concept which has received a lot of attention at JPL recently is the Fizeau Filled-arm Telescope (FFT) mission concept, shown in Figure 2. The FFT is a dilute aperture telescope with crossed 30m x 1m parabolic mirrors which form the image on an array of detectors in much the same way as HST. A team of investigators, led by S. Synnott at JPL, has studied a variety of engineering and science issues related to such a mission. These include structural stability on short (vibrational), moderate (thermal), and long-term (materials) time scales; optics and scientific instrumentation; and science return, both the imaging capability of the instrument and the science possible with that capability.

The advantages of the FFT are

1. it is a filled aperture, in the sense that it samples all points in the pupil plane,
2. it obtains the angular resolution comparable to a filled mirror of many times the weight,
3. it is deployable and can conceivably be collapsed into a volume small enough to be carried in the shuttle or other launch vehicles currently being contemplated,
4. it can perform spectroscopy at the spatial resolution of the instrument in much the

same manner as the HST, if the light is dispersed at an angle skew to the sidelobes.

Figure 3 shows the optical transfer function (OTF) of the 30m FFT and the resulting point spread function (PSF or "dirty beam"). Note the low level plateau in the OTF. For high signal-to-noise sources even these off-diagonal visibilities are sampled by the interferometer, yielding as much spatial information on the source as would be obtained with a filled 30m mirror.

Nevertheless, the PSF has significant sidelobes, compared to a 30m filled mirror. The situation is similar to that for the HST with the current spherical aberration problem. The PSF surface brightness is dominated by a central peak, but most of the energy is in the sidelobes. The FFT, however, has one advantage the aberrated HST does not have. The sidelobes are not even approximately circularly symmetric. A weak source next to a strong stellar object could be detected by simply rotating the FFT until the faint object's position is between the sidelobes. Moreover, one could co-add several images taken at different rotations to obtain an image with sidelobes reduced by a factor of $1/N$, where N is the number of co-added images. Nonetheless, for high quality imaging the weaker sidelobes will still have to be removed by deconvolution, just as has been necessary for ground-based interferometers and is now necessary for HST.

II.   IMAGING TECHNIQUES USED IN SPACE OPTICAL INTERFEROMETER SIMULATIONS AND THEIR APPLICATION TO HST IMAGES

A. Optical Interferometric Imaging Techniques

While space optical interferometric imaging experience has been in the simulated world only (save that of HST itself), our simulations have produced some techniques which may be of general use in any imaging situation requiring deconvolution, including HST. Many of these are specifically related to the CLEAN algorithm, commonly used in radio astronomy. In this algorithm the image is deconvolved by repetitively removing many weak PSFs from the brightest areas in the image until only noise is being removed. The positions and strengths of these PSFs (the so-called delta-function clean components) are tallied and reconvolved with a Gaussian or other smoothing function with full-width-at-half-max equal to that of the PSF main lobe and added to the residuals to yield the final image. While CLEAN may not be the best algorithm for HST-type deconvolution, it certainly is
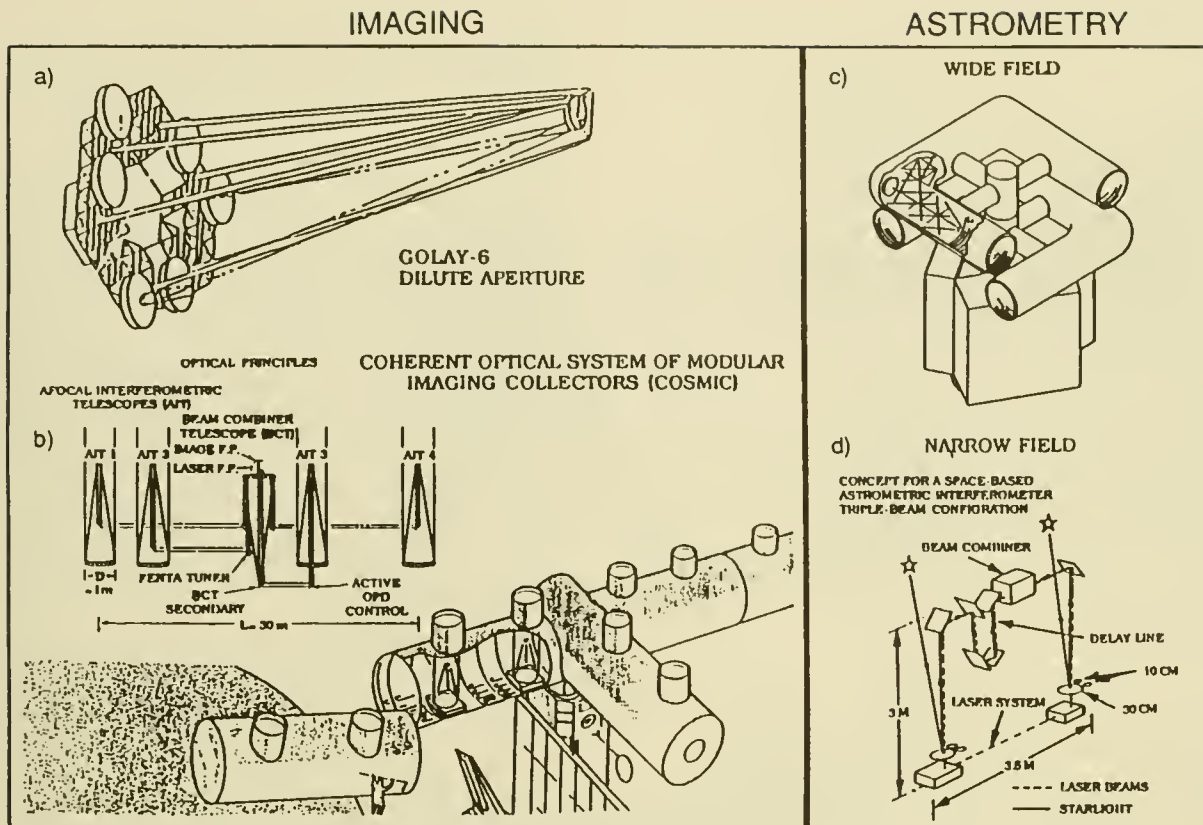
114

Figure 1.   Astrophysics Interferometer Concepts
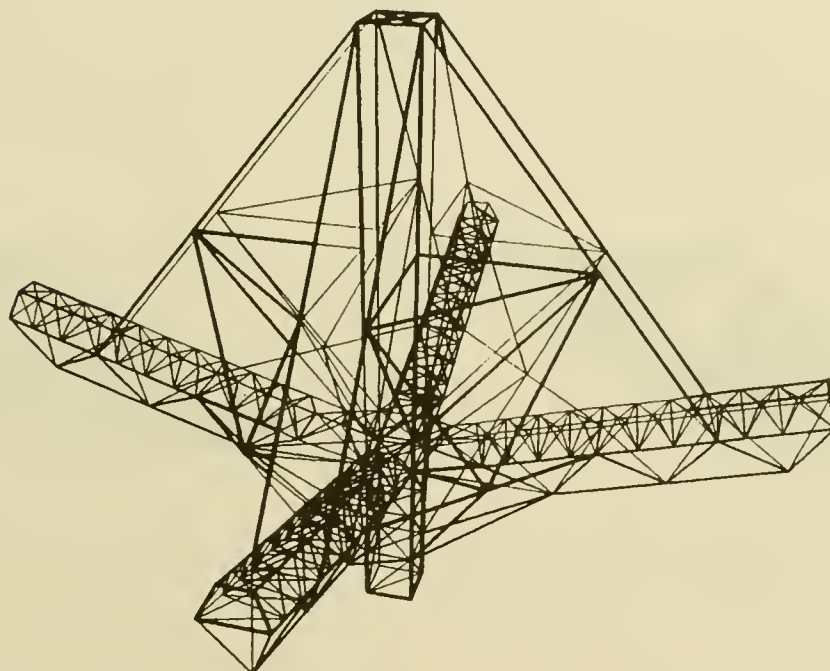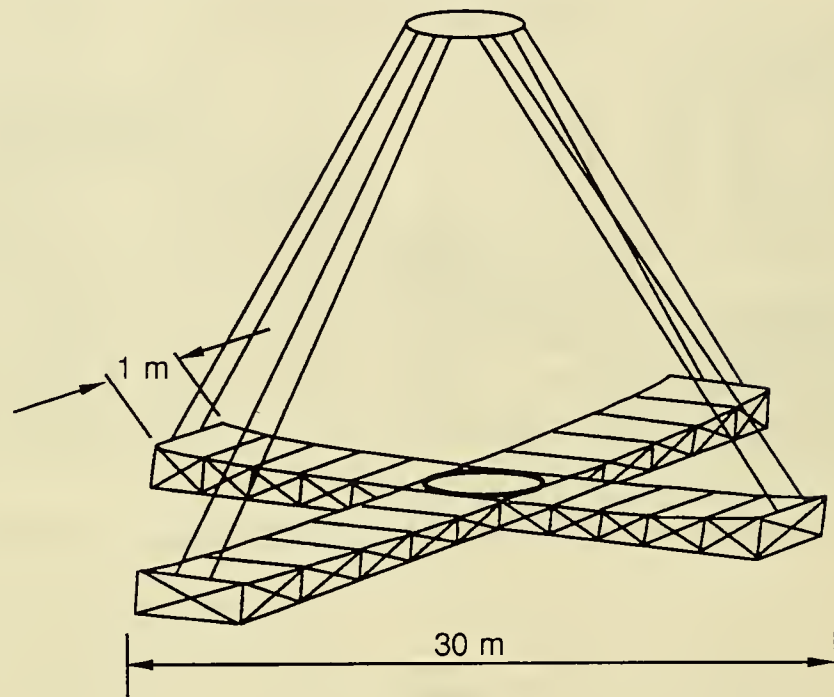

Figure 2.   Fizeau Filled-Arm Telescope

ATTRIBUTES:
- 4 SEGMENTED, PARABOLIC ARMS, 1m x 15m EACH
- TOTAL COLLECTING AREA: 60m$^2 \approx$ 8.7m MIRROR
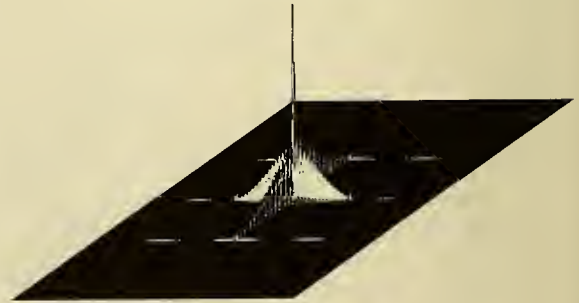

1 m

30 m

SNAPSHOT OTF

SNAPSHOT PSF

Figure 3.   The Fizeau Filled-Arm Telescope

not the failure on optical images that many investigators have claimed. With some moderate image pre-processing CLEAN can be made to work as well on optical images as it has on radio maps.

The first points to be made are trivial, but may possibly be useful to the novice. One should be aware of the units of surface brightness in the various images. The dirty image has units of counts or flux per pixel, whereas the "clean" image (deconvolved with the PSF and restored with a Gaussian or other smoothing function) has units of flux per clean beam effective area. (The "clean beam" is simply the smoothing function.) Some deconvolution software requires the PSF peak to be normalized to unity, rather than the area normalized to unity (one pixel). This gives the PSF a large effective area, in pixels. In order to preserve flux in the deconvolution process, the flux scale of the dirty image should be multiplied by the effective area of the PSF, changing the units to flux per dirty beam area. In addition, one should be careful that the PSF used is centered on the phase center of the pixel array. In the AIPS radio astronomy package, for example, this is at the position (M/2, 1+N/2), where M and N are the X and Y grid sizes, respectively. Other software packages may use a different convention.

It is also obvious, but important, that image deconvolution be the last step in image processing. Effects such as flat fielding, proton hits, and saturated or dead pixels should be taken into account, as the presence of such anomalies during deconvolution can introduce errors worse than the process is trying to remove. Saturated pixels are a particularly severe problem. They cannot be ignored, as dead pixels can, because they result from a very bright source which will cast unsaturated sidelobes into neighboring pixels. These sidelobes must be removed by careful (saturated) PSF fitting because the peak of the bright source is unknown. Care must be taken not to let these sidelobes be mistaken for real sources in the deconvolution process.

We have found two techniques to be particularly useful to the CLEAN algorithm in the deconvolution of optical images: filtering out-of-band noise and subtracting a D.C. floor. CLEAN often fails on well-sampled optical images because of large pixel-to-pixel variations in the flux of sources due to photon statistical noise. This can be overcome easily by filtering out power in the high spatial frequencies, which were never sampled in the first place. (Note that algorithms such as MEM and the Wiener filter automatically filter out such out-of-band noise.) More generally, all Fourier components not specifically sampled by the interferometer should be zeroed out. This is done by thresholding the OTF such that Fourier components with an OTF value less than 10^{-5} - 10^{-3}, say, are set to zero. For HST Planetary Camera images, this amounts to eliminating power on spacings larger than ~2.4m. Wide Field Camera images are undersampled and are likely to not have much out-of-band noise in the first place.

Subtracting the D.C. floor from the dirty image aids the CLEAN algorithm and assists image display software in showing faint features. The floor can be determined theoretically, from known instrument characteristics, or empirically from dark portions of the image. We prefer the latter method as it takes into account a well-resolved source whose size is larger than the field of view (a common occurrence in interferometric imaging), fog due to unfocused mirror segments, sky background noise, readout noise, etc. (Aperture plane interferometers automatically remove the floor if the visibilities measure correlated flux only.) Images with the floor subtracted will commonly have negative features at the noise level. This property can be useful in determining which image features are believable and which are not.

B. Imaging Simulations Using a 30m FFT Instrument

The above imaging techniques were applied to simulated observations of simple and complex objects using a hypothetical 30m FFT instrument. Noise and error sources included in the simulated CCD images were photon statistical noise, Poisson-distributed thermal background noise of 10^{-3} ct s^{-1} pixel^{-1}, Gaussian-distributed readout noise of 1 ct read^{-1}, and a 50% quantum efficiency. The simulations included single integrations (snapshots) of 1000 seconds, as well as "full synthesis" images formed by summing eight such snapshots at different position angles of the instrument. Figure 4 shows one such simulation of a bright 15th magnitude quasar. (All images are displayed with logarithmic contours.) The input model is shown in Figure 4a, with simulated jets, narrow-line region, and bright unresolved objects of absolute magnitude -10 or brighter. The source size is about 0.1 by 0.3 arcseconds.

The dirty image resulting from eight rotations of the 30m FFT interferometer is shown in Figure 4b. Note the long sidelobes emanating from the bright central source and jet knots. A noise floor has been subtracted, and Fourier components not sampled at each rotation have been eliminated, from this image.

Figure 4c shows the CLEANed image after restoring with a 0".006 Gaussian. Note the high fidelity of the image when compared with the original object. Note also the jagged contours on the extended structure, typical of CLEAN results. Compare this image with that produced by a filled mirror with the same amount of glass as is in the 30m FFT (60 m^2 or an 8.7m HST). (See Figure 4d.) Clearly the FFT is a much
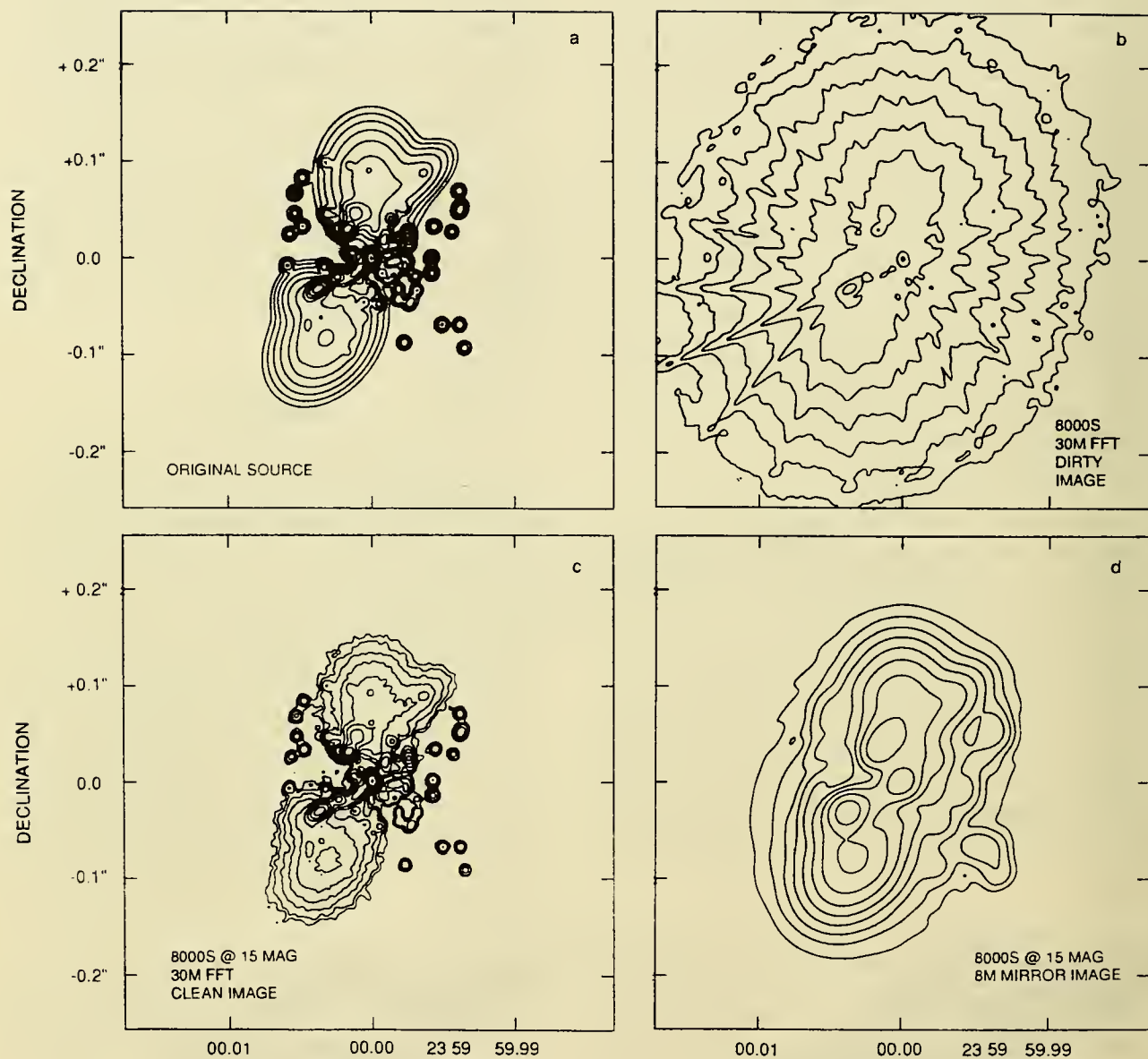
Figure 4.   Simulated Optical Interferometer Observations of
a Model Quasar

better instrument for morphological studies at high resolution.

The 5σ detection limit of the 30m FFT in full synthesis mode is about 27th magnitude when the central peak of the PSF is well sampled (3 x 3 pixels). Science possible with this instrument, in addition to quasar and active galactic nucleus imaging, includes imaging of solar planets (especially Pluto) and their moons, minor planets, some giant and supergiant stars, novae explosions in great detail with high signal-to-noise, and galactic nuclei to search for 10^8 M_o black holes.

C. Suggested General Deconvolution Procedure

In the original talk given at this workshop, the author suggested a hybrid deconvolution procedure based on a radio astronomy technique of CLEANing out bright point sources from a dirty image before using MEM to reconstruct the extended structure. This suggestion was based on CLEAN's well-known poor handling of extended structure and MEM's poor handling of bright point sources. However, this author was sufficiently impressed with some results presented at this meeting -- most notably the quantification of the MEM technique and new resampling procedures -- that he has further modified this suggestion into a procedure for generalized image deconvolution. The final image from this procedure uses the Maximum Entropy Method's ability to handle extended structure, while avoiding its shortcomings with point sources. It also takes maximum advantage of any "superresolved" structure which has a high probability of being correct, while ignoring that which is suspect. This procedure still should be considered suggestive, not definitive, and in need of further investigation. In the interest of saving space it is presented below with little justification. The reader should consult the presentations by Cornwell, Skilling, Nicenson, Wier, Adorf, Lucy, Hanisch, and Horne in this workshop for discussions of QMEM, resampling, and the merits of various algorithms.

1. Fully calibrate and reduce the images before deconvolution, accounting for hot, dead, saturated, or cosmic-ray pixels, bleeding, flat fielding, etc.

2. Filter out high frequency noise, *i.e.*, power in spatial frequencies not sampled. This is necessary for the point source removal step (using CLEAN or DAOPHOT) in the hybrid procedure described below. It would not be necessary if algorithms which automatically reject such power (such as MEM and Wiener filtering) were used exclusively.

3. Remove point sources down to a surface brightness commensurate with the extended structure in the image (not down to zero flux). Use 9-100 point spread functions, rather than one, computed on a finer grid (by a factor of 3-10) than the original image and convolved with a box function of the original pixel size to duplicate the coarser gridding of the dirty image. Subtract out the bright source flux above the extended structure, using CLEAN or DAOPHOT, and record the positions and subtracted flux of the point sources.

4. Deconvolve the residual image with the PSF using the Quantitative Maximum Entropy Method, which assigns an error bar to each pixel in the final image. The final image should also have finer gridding (by a factor of 3-10) than the original dirty image. The comparison dirty image needed by QMEM is generated by convolving with a more finely gridded PSF and a box function to duplicate the coarser gridding of the original dirty image.

5. Smooth the QMEMed image with a smoothing function whose resolution varies across the image, being large where the per-pixel error is large and small where the errors are small. The resulting signal-to-noise of each pixel in the smoothed image should be fairly high (7-10?, I have no quantitative estimate at this time) and approximately uniform across the map. In high signal-to-noise regions of the unsmoothed map, the resolution used could be significantly smaller than the full width at half max of the main lobe of the PSF. This would give a "superresolved" smoothed image in these regions with a high degree of confidence on the superresolved structure.

6. Convolve the point sources removed in Step 3 with the standard FWHM Gaussian and add to the smoothed MEM image. This is the final deconvolved and restored image.

Modifications to this procedure are necessary for each of the three situations we normally consider at JPL: HST, space optical interferometers, and radio interferometry.

a. HST imaging: The HST PSF is not space-invariant, rendering the Fourier

relationship between image and pupil planes inappropriate. In the worst case, this could require storing a different PSF for each pixel in the image -- a four dimensional problem. Less radical approaches divide the image into blocks where the PSF varies slowly and deconvolving with a different PSF in each block.

    b. Space Optical Interferometers: SOIs are likely to have phase errors which do not vary with time. In this case, P. Nicenson's method for deriving both the PSF and image simultaneously should be useful. It is similar in principle to the phase self-calibration technique used in radio interferometry. Note, however, that in general imaging interferometers have partially-filled apertures and therefore _must_ use iterative deconvolution techniques (such as MEM, CLEAN, etc.) instead of Fourier quotient methods like the Wiener filter (which requires filled apertures) which Nicenson uses.

    c. Radio interferometers: Radio interferometers are also plagued with phase errors, but can usually take advantage of phase closure and standard phase self-calibration methods. Resampling can be accomplished even at the dirty image stage, as the detected data are the aperture plane visibilities, not the dirty image. There is then no need for convolving with a box function.

D. Diagnostics of Deconvolution Methods

Evaluation of deconvolution algorithms using simulated HST observations is a very important task which should not end with this workshop. Recently, investigators have been using ground-based images of complex objects (galaxies, star clusters, planets) convolved with the HST PSF to perform the evaluation. These are useful tests, but they are incomplete without simpler tests on individual point sources, to study the effects of sidelobe noise not removed by deconvolution (dynamic range effects), and on close pairs of unequal point sources, to evaluate the photometry near bright stars. (See, for example, T. Lauer's simulations of many close pairs of stars reported

at this workshop.) Several algorithms could possibly be eliminated with such simple tests, and a greater insight into the algorithms could be obtained than with the complex tests.

III. THE IMPACT OF HST DECONVOLUTION EFFORTS ON FUTURE OPTICAL INTERFEROMETRY MISSIONS

The problems of HST image restoration and optical interferometer image restoration are very closely related, especially for missions such as the FFT. We see this workshop as the beginning of an on-going process which will lead to better techniques, not only for HST, but also for future interferometric missions as well.

While the spherical aberration in the HST primary mirror is indeed unfortunate, its discovery has (in a perverse way) created advantages for future missions. First of all, the necessity for HST image restoration will give image deconvolution greater respectability and wider usage in optical astronomy than it currently enjoys. This hopefully will lead to the recognition that unfilled apertures are as important scientifically as filled ones, even though they may emphasize different science. The HST effort will provide the stimulus for in-depth studies of the best optical deconvolution techniques (_cf._, this workshop) and will lead to software which will be directly applicable to space imaging interferometers. The HST effort also will provide practical deconvolution experience in a way simulations could never do, with real detectors and noise sources. The "ground truth" eventually could be provided by WF/PC 2 images. Such experience will help us design better space imaging optical interferometers and to better assess the science possible with such instruments.

Hopefully, the lessons learned from the HST primary mirror problems will lead NASA to establish an on-going optics technology program -- something the optical interferometrists have been requesting for some time. Such a program should help identify optics goals for future missions, establish a number of in-house experts to evaluate contractor work, and provide independent research and testing facilities for space optics. Such a program could be extremely important in ensuring that problems like the HST spherical aberration does not happen again.

Restoration Experiments at the ST-ECF

H.-M. Adorf, J.R. Walsh, R.N. Hook

Space Telescope — European Coordinating Facility
European Southern Observatory
Karl-Schwarzschild-Str. 2
D-8046 Garching b. München
F.R. Germany
Tel.: +49 - 89 - 320 06 - 261

Abstract: Some experiments carried out with four different restoration methods on simulated HST Faint Object Camera f/96 images are described. The restored images were qualitatively and quantitatively evaluated. The two direct linear methods employed, inverse filtering and Wiener-filtering, generally delivered sharper images at the expense of more small-scale artifacts. The two iterative non-linear methods tested, maximum entropy and Lucy-restoration, delivered smoother extended structures with fewer artifacts. To the extent the methods could be tested, none of them restored both point sources and extended structures entirely satisfactorily.

1. INTRODUCTION

The discovery of severe aberration in the optics of the Hubble Space Telescope presents a tremendous challenge to the whole Space Telescope project. Before any fix can be implemented, the optical performance of the telescope and instruments will have to be carefully evaluated, observing proposals will have to be assessed, possibly redesigned, and the precious incoming data will have to be carefully analyzed.

Soon after it became clear that image restoration would play a vital role in this endeavour, the ST-ECF decided to carry out some experiments in order to become familiar with the techniques suited to the aberrated HST point-spread function (PSF). At the time when the experiments were initiated (early August), the Science Assessment / Early Release Observations were not yet available. HST was still outgassing and a final focus position had not been agreed upon. We therefore decided to apply several readily available restoration methods to realistic simulated images.

2. SIMULATIONS

We wanted to carry out restoration experiments on a simulated Faint Object Camera f/96 field. The FOC, for which an observed point-spread function was available, was modelled as an ideal photon counting device with a space-invariant PSF. From an encircled energy plot of the observed PSF an artificial, circularly-symmetric PSF was derived, capturing the sharp central peak and the extended halo structure of the observed PSF, but not the fine details such as the "hairy tendrils". For comparison we also employed an artificial nominal PSF in our experiments.

We chose to simulate a "difficult" field containing a number of point sources on top of an extended source. To this end, at each of 300 star positions from the Pleiades field a suitably resampled aberrated PSF was inserted into a 512x512 frame (0.022 arcsec pixels). Intensities were scaled such that the brightest star had a count rate of 1 Hz. The dynamic brightness range was 7.5 mag. The star field was co-added to a heavily smoothed (stars removed) image of the galaxy M101, convolved with the artificial PSF. Finally, a background of 3×10^{-3} counts/pixel/sec was added and the image was powdered with Poisson noise.

Three simulated observations of this artificial input image were prepared, two with the artificial aberrated PSF and one with the nominal PSF. The aberrated simulations were carried out for 1,000 sec (a typical HST exposure time) and 30,000 sec (a fairly unrealistic exposure time). The simulation with the nominal PSF was done for 1000 sec. For reference purposes we also kept a noise-free nominal PSF image.

With the wisdom of hindsight we note a few shortcomings of our experiments: Firstly, we missed out on the true two-dimensional nature of the real PSF with its peculiar power-spectrum (see di Serego Alighieri & Walsh 1990). In view of the importance of high frequencies for restoration, the absence of much of the PSF fine-structure casts some doubts on the validity of our experiments. Secondly, the restoration algorithms used the same noise-free PSF as the simulations generating the data, leading to systemati-

cally over-optimistic restoration results. Instead, we should have used several different noisy PSFs enabling us to carry out a sensitivity analysis.

3. RESTORATION METHODS AND THEIR IMPLEMENTATION

Four restoration methods were considered in our experiments, namely linear "inverse filtering" and "Wiener filtering", and non-linear "maximum entropy" and "Lucy-restoration". The fast linear filter methods were implemented with a view towards using their results as initial estimates for subsequent iterative restorations. Lucy's method was chosen since the original author was present at ESO, maximum entropy for its code-availability and prevalence in radio-astronomy. Note, however, that none of the methods is based on realistic imaging equations.

Other restoration methods have been tested outside the ST-ECF using our simulated HST data: a "constrained maximum entropy method" (Reiter & Pfleiderer 1986; Reiter 1990) and the "minimum information method" (Pfleiderer 1985).

3.1. Modified inverse filtering

The standard inverse filter restoration method (see e.g. Andrews & Hunt 1977, p. 148), sometimes called Fourier-quotient method, carries out an unconstrained algebraic inversion of a linear, deterministic, space-invariant imaging equation. The method was implemented on the procedure level in both the MIDAS and the Interactive Data Language (IDL) image processing systems. Inverse filtering requires 2 forward fast Fourier-transforms (FFTs) and 1 backward FFT and, due to its non-iterative nature, is relatively fast.

The direct application of inverse filtering, the simplest of all restoration methods, to our simulations produced a considerable number of small-scale artifacts. An obvious remedy to this problem is to replace the goal of a *perfect restoration*, characterized by a δ-function, by a *finite resolution restoration*. When a low-pass filter — we employed Gaussians with a σ of 30 to 50 frequency bins — was added to the basic method, a large number of those artifacts could be suppressed.

3.2. Wiener-filtering

The idea behind a low-pass filtered Fourier-quotient method is carried further by Wiener-filtering (see e.g. Andrews & Hunt 1977, p. 150). This method explicitly acknowledges the presence of (Gaussian additive) noise in the data and replaces the *experimentally* determined "best" low-pass filter by one which is *designed* to optimize the signal-to-noise ratio on the basis of signal and noise power-spectra estimates. However, since the standard Wiener-filter method adapts to the *global* rather than to the *local* S/N in the data frame, some regions of interest may become under-restored, others over-restored, when the S/N ratio varies substantially across the data frame.

We implemented the Wiener-filter method in IDL. The required signal power-spectrum estimate was derived interactively from the distorted data frame by smoothing its power-spectrum with a Gaussian of σ=5 frequency bins and padding the high-frequency (noise-only) regions with zeros. The noise power-spectrum, assumed to be white, was set to a constant fitted to the high-frequency power. The necessity of manual intervention for filter design is a shortcoming of our current IDL-implementation of the Wiener-filter. Apart from filter design, Wiener-filtering has the same computational requirements as modified inverse filtering.

3.3. Lucy-restoration

Lucy's recursive restoration scheme (Lucy 1974; Lucy & Baade 1989; Baade & Lucy 1990) appeared to be an attractive restoration method for several reasons: Firstly, it is simple to implement. Secondly, it is theoretically sound, obeying e.g. the non-negativity constraint of optical astronomical images and preserving flux *locally*. Thirdly, during the iterations higher frequencies are only gradually enhanced. Given that it is not advisable to restore to convergence, an obvious shortcoming of the basic iteration scheme is that it does not come with a built-in stopping rule.

The existing general MIDAS-implementation of Lucy's method works exclusively in image space and is impractical for the FOC-restoration problem with its large frame sizes and a large, space-invariant PSF. We therefore re-implemented Lucy's method restricted to space-invariant PSFs in an IDL-procedure exploiting fast FFTs. In each iteration 2 forward and 2 backward FFTs are computed. For the required initial estimate for the restored image we usually employed a constant mean-intensity frame.

We also experimented with a Wiener-filtered image as initial estimate in order to save iteration cycles. However, since the iteration process has to be stopped before convergence is reached, the result undesirably depends to some extent on the history of the process. A linearly filtered initial estimate tends to introduce high frequencies right at the beginning, which Lucy's method, starting from a constant initial estimate, would bring in only gradually. Furthermore, a linear space-invariant method almost invariably introduces meaningless negative values into the initial estimate, preventing an exploitation of the non-negativity

property of real images. The idea of using a non-flat initial estimate for Lucy's method was therefore abandoned for the time being.

3.4. Maximum entropy
Experiments with the maximum entropy method (MEM, see e.g. Bryan & Skilling 1980) were performed using a fairly outdated Starlink-implementation from 1983, now superseded. However, the method and any comparisons drawn from it are useful. The MEM-program requires as input the observed/simulated data frame, a (non-critical) estimate of the total signal of the observed frame, an error image containing the inverse of the pixel-by-pixel variance of the signal in the data frame or, alternatively, a constant noise standard deviation per pixel, the point spread function, an initial estimate of the restored image and finally the number of iterations.

4. ANALYSIS OF THE RESTORED IMAGES
Table 1 shows the same subarea of our simulated FOC f/96 frame in various unrestored and restored forms. Any comparison of the different results is driven by questions such as: Which method performs better on low S/N data? Is there a method which simultaneously optimally restores point sources *and* extended structures? Can one quantitatively work on restored images? Which method is more suitable to photometric, which more to astrometric measurements?

We have visually inspected our restoration results and summarize our findings below. In addition to this qualitative assessment we attempted a quantitative comparison of the methods by measuring several parameters on the restored images: 1. "sharpness" defined through the encircled flux of point sources, 2. "positional fidelity" through the x- and y-centroids of point sources, 3. "photometric fidelity" through aperture flux measurements of point sources and 4. image statistics through mean and standard deviation of the intensities of extended structures.

4.1. Linear restorations
It came as a surprise that the Fourier-quotient method followed by a low-pass filter worked so well on our simulated FOC f/96 data. However, the background and the galaxy showed some mottling. Also, there remains some arbitrariness in the characteristics (i.e. shape and frequency cut-off) of the low-pass filter and the determination of a visually optimum one requires experimentation. Nevertheless, if sharp images is the primary restoration goal and the S/N is sufficiently high, modified inverse filtering is certainly worth a consideration for HST image restoration.

Results from Wiener-filtering appeared not to be substantially better than those obtained with modified inverse filtering. The galaxy and background also appeared speckled, particularly on the low S/N restored image, although there were fewer small-scale artifacts. Some ringing around point sources, characteristic of linear restoration methods, could also be seen.

4.2. Non-linear restorations
Generally speaking we were quite satisfied with the results obtainable with Lucy's recursive restoration scheme. Extended structures were well restored; point-sources, on the other hand, appeared somewhat less sharp when compared to the linear methods and we saw also some ringing around point sources on non-zero background. The price to pay are the many iterations (about 40 in our case) attributable to the slow convergence of the simple Picard fixed-point iteration algorithm used. As a stopping criterion for the iterations we used the visual appearance, although in the future a quantitative divergence measure (e.g. the reduced χ^2) will be used instead.

The MEM-program was run for 20 iterations with error maps derived from the observed frames, assuming Poisson statistics. The results after 20 iterations showed more structure than those after only 10 iterations. The inital estimate was taken identical to the original image. A run in which this estimate was a flat frame gave a markedly inferior result after the same number of iterations.

5. DISCUSSION
Our restoration experiments yielded several results, some objective and some subjective. We implemented three simple restoration methods (modified inverse filter, Wiener-filter and Lucy-restoration), using two different image processing systems (MIDAS and IDL), and developed a feeling for their CPU-time and storage requirements. The other tested standard restoration method (maximum entropy) surprised us by not handling point-sources too well. In applying these algorithms to our simulations we gained experience in practical restoration work, and a feeling for how well simple methods are suited to the HST-type PSFs and what kind of artifacts they produce.

Generally speaking the results from the linear and the non-linear methods fall in two separate groups: the two linear methods deliver sharper images, whereas the two non-linear methods deliver smoother extended structures and fewer artifacts. So far, none of the methods restores both point sources and extended structures entirely satisfactorily. Photometry and astrometry both seem possible on restored images. No method turned out to be astrometrically much superior to the others; Lucy's recursive scheme performed well, MEM somewhat less so.

The comparison of the restorations from data frames with high and moderate S/N reveals a marked difference between the linear and non-linear methods: The *high S/N* frames (not shown) are restored quite well by all the methods tested. However, the methods behave quite distinctly on the *moderate S/N* frames (Fig. 1). The modified inverse filter (Fig. 1c) delivers high resolution of point sources at the expense of random speckles everywhere. The Wiener-restoration (Fig. 1d) shows essentially the same resolution, but displays some distracting regular texture (of unknown origin). The Lucy-scheme after 40 iterations (Fig. 1e) restores much of the extended structure of the galaxy without introducing obvious artifacts. Going to 80 iterations (not shown) improves the resolution, but also starts to produce speckles at the locations of noise spikes. The maximum entropy method after 40 iterations, much to our surprise, did very little to the distorted image. (It also did not introduce artifacts.) The reason for this "conservative" behaviour is presumably the predominance of the smoothing constraint.

We stress that the preliminary assessment above is based on results obtained with a "naïve" application of the restoration algorithms to simulated data. Whereas we see little room for improvements of the linear methods, we believe that the non-linear methods have potential, particularly for moderate S/N data, which we have not yet exploited.

Our experiments made us aware of the abundance of convolutions in restoration work, stressing the importance of compute- and storage-efficient FFT-implementations not restricted to dataset sizes of powers of two. The image processing systems used (IDL, MIDAS and later STSDAS) provided a convenient environment for quick implementations of restoration algorithms on the procedure level. However, we found that only IDL procedures are really suited to the compute-intensive iterative restoration algorithms, by avoiding frequent disk-accesses for storing intermediate results. Elapsed times for pure FFTs also varied considerably between the current implementations in the three image processing systems used.

The choice of the restoration method appears somewhat less important than anticipated, although better restorations (rated by displaying fewer artifacts) are clearly obtained with the compute-intensive non-linear methods. It can be expected that the methods ultimately to be used for HST-restoration will be an amalgam of those restoration methods, routinely applied in radio-astronomical aperture-synthesis interferometry, and statistical estimation methods used in X-ray astronomy. The radio-astronomy example teaches us that it may take years before algorithms mature and opinions on their utility and reliability converge.

6. OUTLOOK

We consider the experiments described here to be only a first step in a series of actions the Space Telescope Science Institute and the Space Telescope — European Coordinating Facility are expected to jointly undertake in the area of HST image and spectra restoration. A problem obviously deserving attention is the derivation of method-specific error maps allowing an objective assessment of the reality of features appearing on restored images. The derivation of the method-independent statistical Cramér-Rao error bounds, which cannot be improved upon by any method, seems also feasible, provided that the stochastic models of the instruments are available in a readily accessible form.

To optimize existing algorithms, to generalize and adapt them to the specific needs of HST seems an obvious goal, too. Finite resolution restoration, rational stopping rules for iterative restoration schemes and subsampling options are cases in point. The severe undersampling of the Wide Field and Planetary Camera (Adorf 1989a, b) together with its strongly space-variant PSF will present a challenge to any prospective image restorer.

The ST-ECF plans to test a more recent implementation of the maximum entropy method and to compare the computing and storage requirements of various algorithms. We intend to monitor progress on more comprehensive statistical restoration methods such as maximum likelihood and Bayesian methods. Finally, we intend to collect implementations of proven restoration algorithms, and to package and re-distribute them on request to the European HST community.

POSTSCRIPT

Immediately after the workshop two of us (HMA and RNH) had the opportunity to work extensively with real data from the HST Science Assessment Observations campaign. This experience shifted focus away from the different methods to data pre-processing and representation problems. We became aware of the importance of the effects produced by the real cameras, e.g. the non-linearity and geometric distortions of the FOC. We also realized that the various restoration methods display quite a different level of robustness with respect to PSF-errors.

ACKNOWLEDGEMENTS

REFERENCES

Adorf, H.-M.: 1989a, "Coping with the undersampling of WFPC images", *ST-ECF Newsl.* **12**, 9—10

Adorf, H.-M.: 1989b, "On the HST Wide Field and Planetary Camera Undersampling Problem", in: *Proc. 1st ESO/ST-ECF Data Analysis Workshop, Garching, April 1989*, P. Grosbøl, R.H. Warmels, F. Murtagh (eds.), pp. 215—218

Andrews, H.C., Hunt, B.R.: 1977, "Digital Image Restoration", Prentice-Hall International, Inc., London

Baade, D., Lucy, L.B.: 1990, "HST images: What can image processing do?", *Messenger* **61**, 24—27

Bryan, R.K., Skilling, J.: 1980, "Deconvolution by maximum entropy, as illustrated by application to the jet of M87", *Monthly Not. Royal Astron. Soc.* **191**, 69—79

di Serego Alighieri, S., Walsh, J.R.: 1990, "The HST Point Spread Function", *ST-ECF Newsletter* **14**, 4—7

IDL User's Guide, Interactive Data Language Version 2.0, March 1990, Research Systems Inc., 777 29th Street, Suite 302, Boulder, CO 80303, USA

Lucy, L.B.: 1974, "An iterative technique for the rectification of observed distributions", *Astron. J.* **79**, 745—754

Lucy, L.B.: 1990, "Deconvolution with resampling — images and spectra", in: *Proc. Workshop "The Restoration of HST Images and Spectra"*, Space Telescope Science Institute, Baltimore, 21.—22. Aug. 1990, (this volume)

Lucy, L.B., Baade, D.: 1989, "Deconvolution with Increased Sampling", in: *Proc. 1st ESO/ST-ECF Data Analysis Workshop, Garching, April 1989*, P. Grosbøl, R.H. Warmels, F. Murtagh (eds.), pp. 219—223

Pfleiderer, J.: 1985, "Minimum Information Method: Image sharpening by minimization of structural information", *South African J. of Physics* **8**, 73—78

Reiter, J., Pfleiderer, J.: 1986, "Improvement of MEM-deconvolution by an additional constraint", *Astron. Astrophys.* **166**, 381—392

Reiter, J.: 1990, "An algorithm for deconvolution by the maximum entropy method with astronomical applications", *Schwerpunktprogramm der Deutschen Forschungsgemeinschaft "Anwendungsbezogenen Optimierung und Steurerung"*, Report No. 229, 30 pp

Figure 1: A subarea of the simulated, moderate S/N (1000 sec), FOC f/96 frame in various distorted
(next page) and restored forms. (a) Simulated observation with an artificial, circularly-symmetric "nominal" PSF; (b) simulated observation with an artificial, circularly-symmetric "aberrated" PSF; (c) modified (i.e. low-pass filtered) inverse-filter restoration; (d) Wiener-restoration; (e) Lucy-restoration after 40 iterations; (f) maximum entropy restoration after 40 iterations.
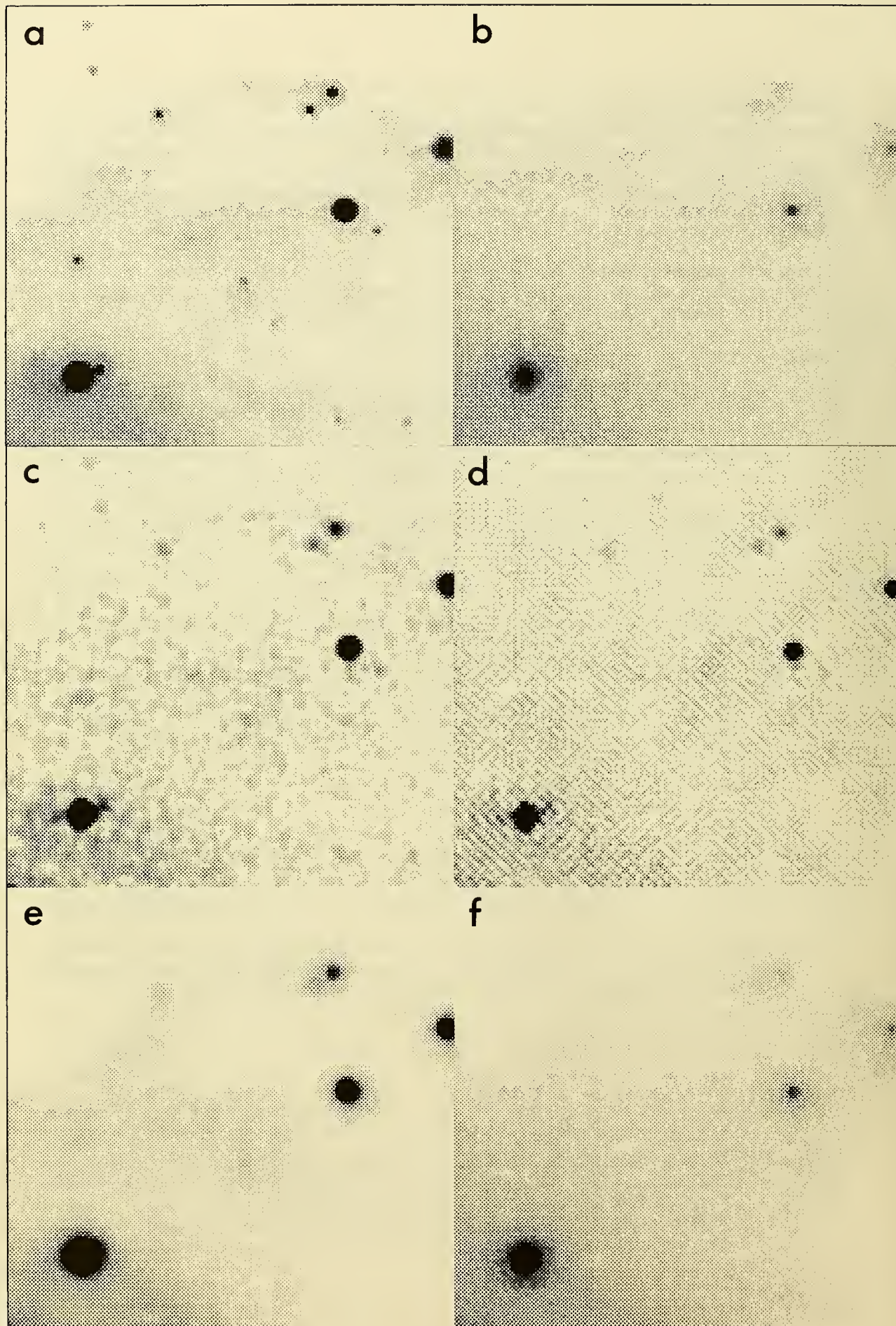
Linear Restoration Techniques:
Not All Bad

Robert J. Hanisch
Space Telescope Science Institute
3700 San Martin Drive
Baltimore, MD 21218

Abstract

Linear image deconvolution techniques such as the Fourier inverse and Wiener filter are often rejected for use in high fidelity image restoration work because of several types of artifacts that can be induced in the restored data. However, for certain scientific objectives the linear techniques can provide acceptable, and in some cases, preferable results. In this paper the Fourier inverse and Wiener filter techniques are reviewed, and sample restorations of HST images from both the Wide Field/Planetary Camera and the Faint Object Camera are shown.

Introduction. Many workers in the image restoration field have pointed out the difficulties in using linear restoration techniques such as the Fourier inverse and Wiener (optimal) filter. The primary objections to these techniques are that they 1) amplify the noise and 2) introduce artifacts at high spatial frequencies (i.e., ringing), making interpretation of the restored images difficult. Subtle morphological features are not likely to be recovered using simple linear image restoration techniques. However, one clear advantage of the linear techniques (as the name implies) is that the relative brightnesses of objects in the field are preserved. Moreover, the linear restoration techniques are computational efficient. Therefore, for certain classes of image restoration problems these techniques may be more than adequate.

As an example, consider the problem of doing crowded field stellar photometry on HST images. The extended point spread function (PSF) makes object identification difficult. A linear restoration of the image could be used to determine the initial object list, with no corruption of the relative brightness measurements. Overlapping PSFs make it very difficult to obtain a good estimate of the sky background, which compromises the quality of the photometry obtained from PSF fitting (see, for example, Holtzman et al. 1991). Much testing and experimentation is required in order to determine what combination of PSF fitting and deconvolution is optimal for crowded field photometry.

Linear Restoration Techniques. The two techniques investigated here are the standard Fourier inverse and the Wiener (optimal) filter. Although these techniques are well-documented in the literature (see, for example, Ekstrom 1984), for the reader's convenience a brief summary is given below.

In image restoration problems one assumes that the observed brightness distribution $o(x, y)$ can be expressed as the convolution of the actual brightness distribution $a(x', y')$ with the point spread function $p(x - x', y - y')$. In addition, the convolved brightness distribution is contaminated by a component of random noise, $n(x, y)$. Thus,

$$o(x, y) = n(x, y) + \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p(x - x', y - y')a(x', y')dx'dy' \tag{1}$$

If this is rewritten in terms of the Fourier transforms O, N, P, and A of the functions o, n, p, and a, one obtains

$$O(u, v) = N(u, v) + P(u, v)A(u, v) \tag{2}$$

since a multiplication in the Fourier domain is equivalent to a convolution in the image domain.

The Fourier transform of the deconvolved image A' is found by simply dividing the function O by the Fourier transform of the PSF:

$$A'(u, v) = \frac{O(u, v)}{P(u, v)} = A(u, v) + \frac{N(u, v)}{P(u, v)} \tag{3}$$
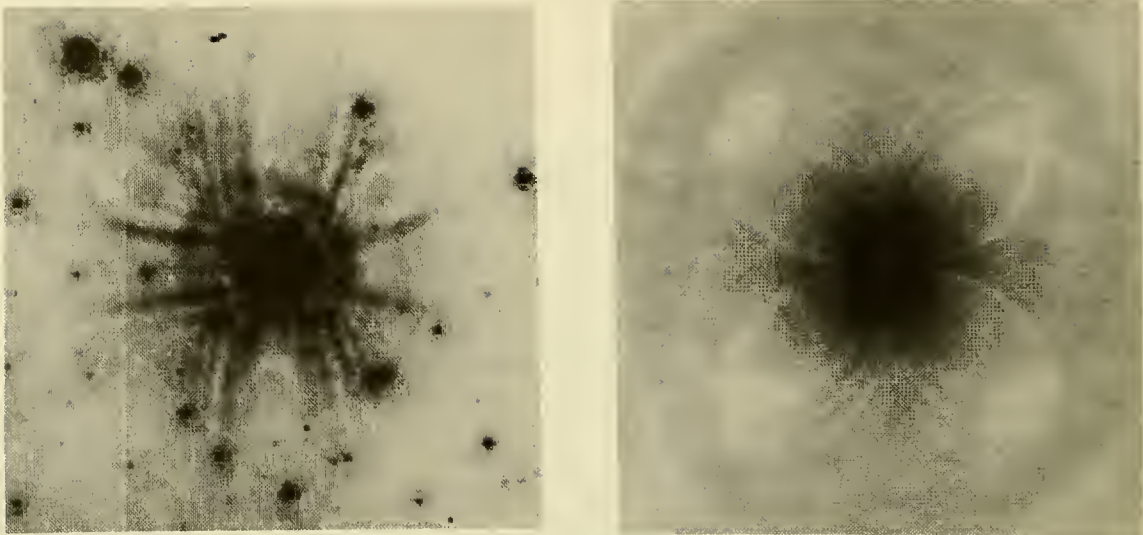
127

Figure 1: *On the left is a typical HST point spread function (this is from the Planetary Camera in filter F547M). The linear gray scale has been stretched to show the structure in the wings of the PSF. Neighboring faint stars and cosmic ray hits have not been removed. The image on the right is the amplitude of the Fourier transform of the PSF shown on the left, after cosmetic removal of faint stars and cosmic ray hits.*

As Eq. 3 indicates, in any region where the Fourier transform of the PSF is small or zero, the noise term in the equation will dominate and the noise is amplified. In the radio interferometry case, where P is not fully sampled, the Fourier inverse technique cannot even be used. Because HST has a filled aperture, the Fourier transform of the PSF tends to be well-behaved with few zeroes (see Fig. 1).

The restored image $a'(x, y)$ is obtained by computing the Fourier transform of $A'(u, v)$. This image, however, is restored to "infinite" resolution, i.e., the PSF has been totally removed from the image, and star images appear as delta functions. This is aesthetically not the most pleasing image, and typically one convolves a' with a smoothing function (i.e., a gaussian of FWHM equal to the that of the core of the unaberrated PSF) for display and analysis purposes. This smoothing is also used to reduce the noise induced into the restored image appearing at the Nyquist frequency. The same objective can be obtained by dividing not by the Fourier transform of the PSF in Eq. 3, but by the Fourier transform of the function which restores the nominal HST resolution. This function is simply P divided by a gaussian of the appropriate width.

The Wiener filter is a modification of the Fourier inverse which reduces the noise amplification problem of the latter technique. The Wiener filter is derived by minimizing the differences, in the least squares sense, between the restored image $a'(x, y)$ and the observed image $a(x, y)$. The Wiener filter P_w is given by

$$P_w(u, v) = \frac{P(u, v)^*}{|P(u, v)|^2 + \frac{S_n(u,v)}{S_a(u,v)}} \tag{4}$$

where P^* is the complex conjugate of P, S_n is the power spectrum of the noise, and S_a is the power spectrum of the signal. This function may be inserted into Eq. 3 in place of P, or the Fourier transform of P_w may be used to reconstruct the image a':

$$a'(x, y) = p_w(x, y) * o(x, y) \tag{5}$$

Application to HST Images. The Fourier quotient and Wiener filter techniques have been applied to a variety of HST images, and some sample results are shown in Figs. 2-4. Prior to deconvolution the images have had any defects (such as data drop-outs or cosmic ray hits) repaired by interpolating over the affected pixels. This is an essential step in the use of the linear algorithms, since any sharp feature in the image which does not exhibit the characteristics of the PSF will induce a ghost image of the PSF into the restored image.

Compared to non-linear, iterative restoration techniques, the linear techniques are more sensitive to having a good representation of the PSF. Because of the strong position dependence of the PSF in the Wide Field/Planetary Camera (which results from vignetting in the WFPC's internal repeater optics), the linear

techniques cannot be used to reconstruct an entire WFPC frame using a single PSF model. Fig. 2 shows the results of deconvolving a PC image of the R136 region using three different models of the PSF based on three different stars in the field. The spatial variance of the PSF is immediately obvious: only a region of about 50 pixels radius around the PSF star is restored acceptably. The mismatch between the model and actual PSF in other regions of the detector leads to strong PSF residuals in the restored images. These data indicate that high quality linear restorations will require an image to be broken up into perhaps an 8 × 8 or 10 × 10 grid of subimages, each with a separately determined PSF.

The Faint Object Camera does not have a spatially variable PSF, but is more susceptible to saturation. In addition, for the high f-ratios the PSF extends over essentially the entire field of view, potentially making it very difficult to correct for the light contributed from stars just outside the image.

Fig. 3 shows a Fourier quotient restoration of SN 1987A as observed with the FOC. The PSF was constructed by combining the images of the two neighboring stars, carefully matching the positions and flux levels to make one PSF which is not contaminated by light from the supernova or its surrounding ring. The restored image shows much improved resolution of the ring – in many regions the ring is unresolved in the transverse direction, and the brightness level interior to the ring is not significantly different from that outside the ring, indicating that the ring is not a result of simple limb-brightening through a spherical distribution of ionized gas. That is, the ring is, at least in part, truly an annulus. Isophote fitting to the deconvolved image shows that an elliptical isophote fits the observed image of the ring very well.

The Wiener filter was used to construct a restored image of Saturn as observed with the Wide Field camera. The PSF for this image was taken from another image in the Science Assessment/Early Release Observations program. The restored image (Fig. 4) shows many details lost in the original data: the atmospheric bands are well-defined, and the Encke division near the edge of the rings is clearly resolved. The image of Saturn's disk, however, contains some undesirable high-frequency artifacts. While many surface features of Saturn can probably be identified and tracked successfully with just a Wiener filter deconvolution, detailed morphological studies would clearly benefit from more sophisticated restoration algorithms.

Both the Fourier quotient and Wiener filter algorithms are computationally very efficient. On a Sun 4/490 fileserver (rated at approximately 25 MIPS and 4 MFLOPS) the Fourier quotient requires about 30 CPU seconds to process a 512 × 512 pixel image with single precision floating point pixel values. The computations include two FFTs (forward transform of the image, forward transform of the PSF), a complex divide, a third FFT (inverse transform of the quotient), and a post-processing low-pass filter. The Wiener filter computations for the same size image take only slightly longer.

Conclusions. Although linear image deconvolution algorithms have serious limitations, their computational efficiency and simplicity of use, combined with the relatively good behavior of the Fourier transform of the HST PSF, makes such techniques both applicable and useful for HST image restorations. These techniques may prove to be useful in determining initial flux estimates for PSF fitting in crowded fields, and for providing a quick initial image restoration to help assess the utility of using more sophisticated (and time consuming) algorithms.

References

Ekstrom, M. P. (ed.) 1984. *Digital Image Processing Techniques*, Academic Press (Orlando).
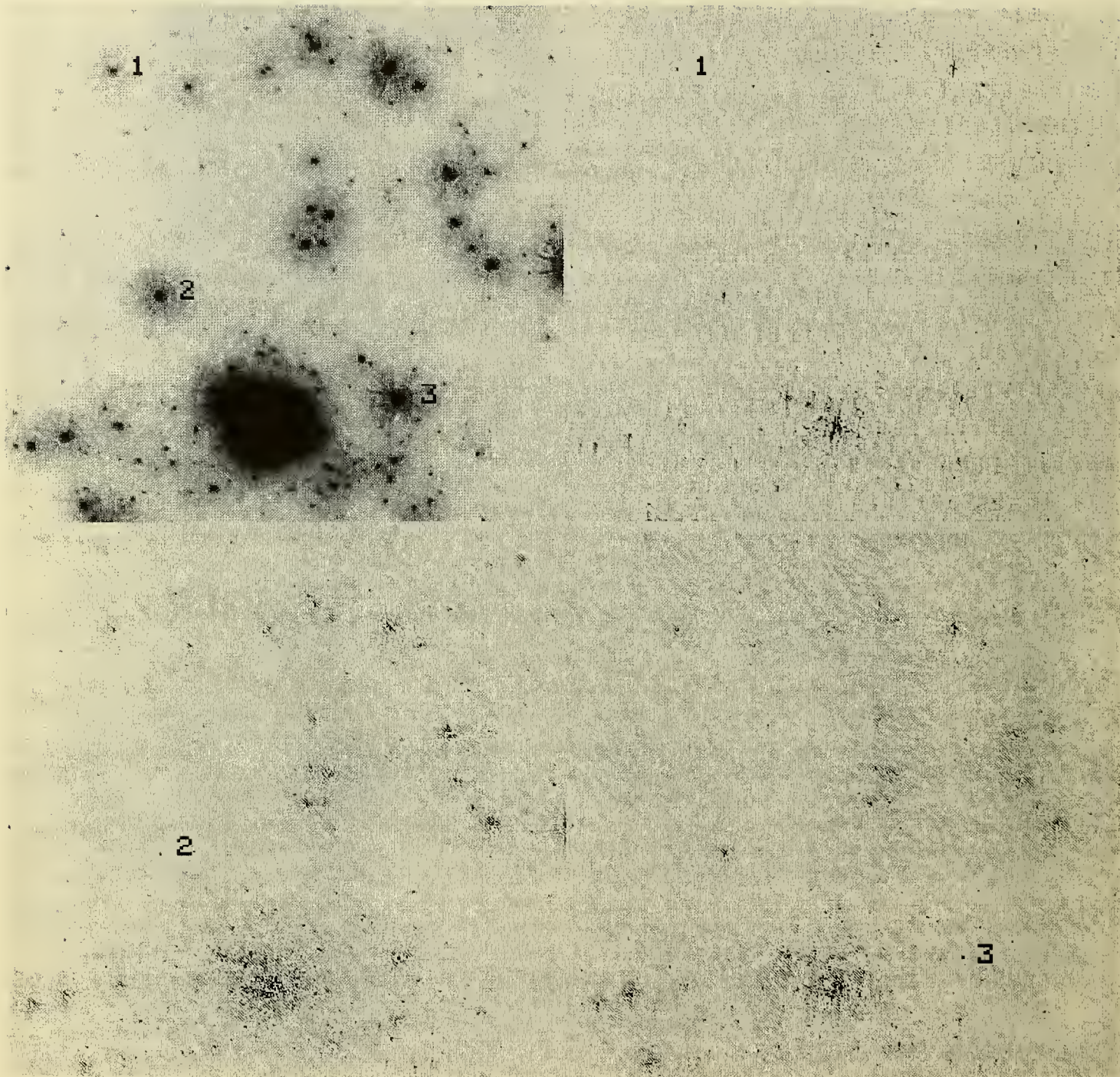
Holtzman, J.A., et al. 1991. ApJ Lett, in press.

Figure 2: *The spatial variability of the PSF is demonstrated in these three Fourier quotient deconvolutions of a PC image of the star cluster R136. The original image is shown at upper left, with three model PSF stars labelled '1', '2', and '3'. The image at upper right shows the deconvolution as computed using star 1 as the PSF, the image at lower left shows the deconvolution resulting from star 2 as the PSF, and the image at lower right results from star 3 as the PSF. Only in the immediate region of the PSF star is each image deconvolved correctly. In regions only 100 or so pixels distant, the artifacts in the deconvolved image show a clear mismatch between the model and true PSF. High fidelity image restorations will require having a grid of perhaps 100 model PSFs for each of the four WF and each of the four PC CCDs for each filter.*
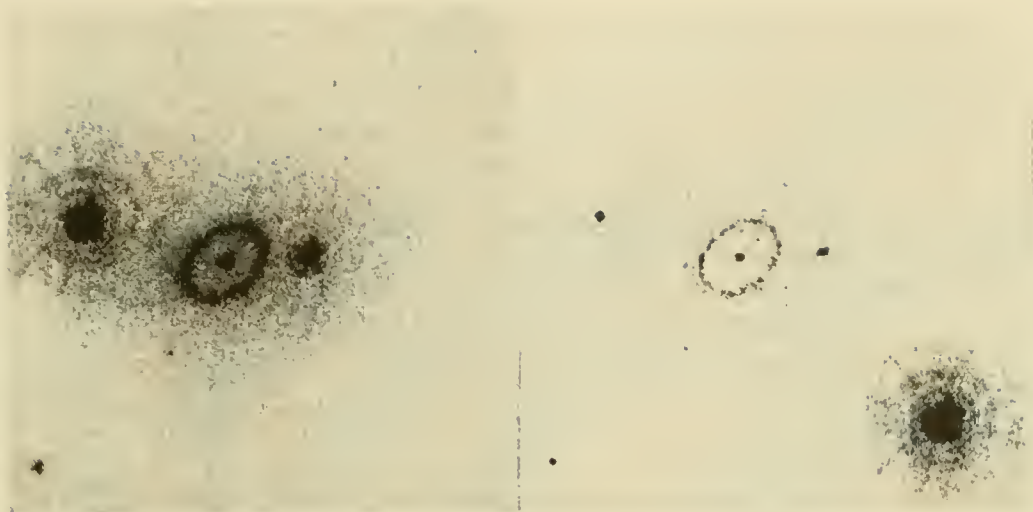
Figure 3: *An FOC image of SN 1987A is shown at left, and the reconstruction of the image using a direct Fourier quotient is shown on the right. The PSF for this field (shown in the inset at lower right) was constructed from the two neighboring stars. A slight discontinuity in the restored image just to the left of the ring is caused by a somewhat abrupt cutoff at the edge of the PSF model; further apodization of the PSF would probably diminish the magnitude of the artifact.*
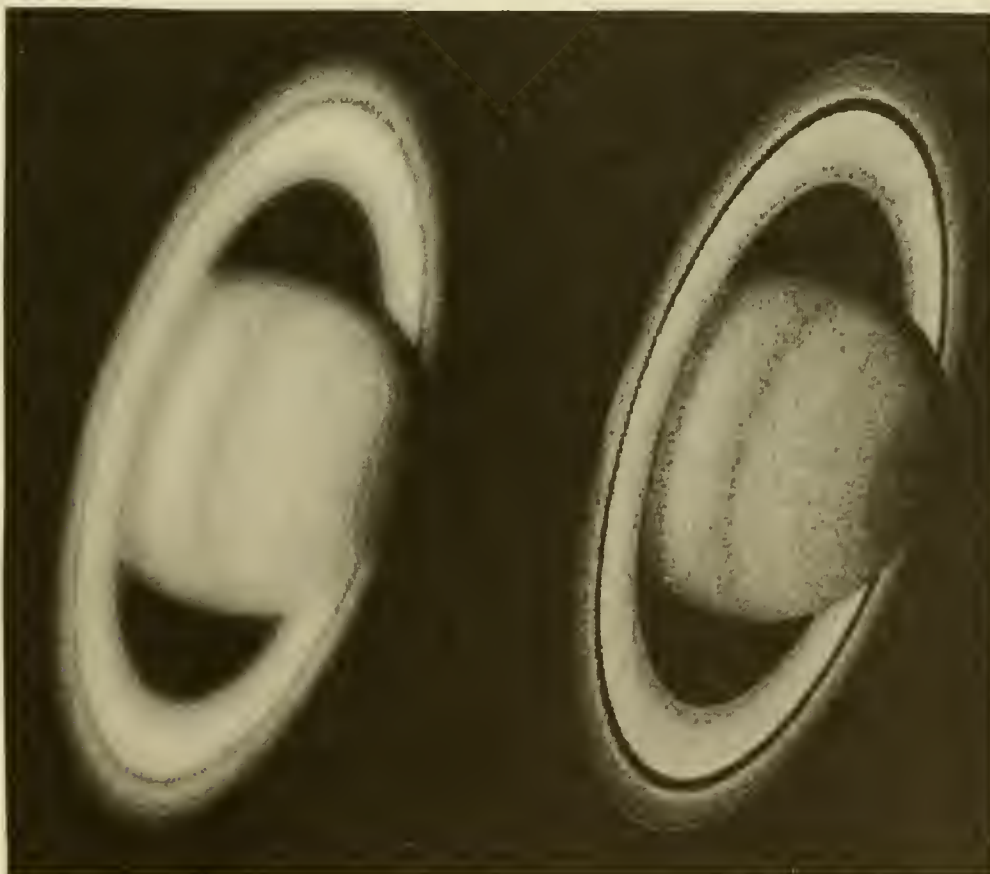

Figure 4: *A Wide Field Camera image of Saturn appears on the left with its Wiener filter reconstruction shown on the right. The Wiener filter image is a reasonably good reconstruction, clearly showing the Encke division in the rings and the band structure in Saturn's atmosphere. The mottled structure visible in the disk is typical of the type of high frequency artifacts inherent in the Wiener technique. Such effects are considerably worse in the Fourier inverse reconstruction of this image.*

Maximum Entropy Deconvolution
of a Wide Field Camera Image of R136
Keith Horne, STScI

1 Summary

A manufacturing error has introduced spherical aberration in the primary mirror of the Hubble Space Telescope (HST). The resulting point-spread function (PSF) extends in radius to several arcseconds, but with significant diffraction-limited structure. In particular, about 1/5 of the light is concentrated in a sharp core. Thus while detection sensitivity is degraded by a factor of order 5, deconvolution methods can be used to recover nearly diffraction-limited imaging for suitably bright targets.

We present results of a test using the maximum entropy method (MEM) to deconvolve an image of R136, the central star cluster of the 30 Doradus nebula, from data taken with HST's Wide-Field Camera (WFC). The deconvolved image reveals hundreds of stars within a radius of 2 arcseconds.

MEM's flexibility allows us to fit the image to the CCD data through a fairly realistic model of the detector and its noise characteristics. For example, while the 0.1 arcsecond WFC pixels undersample the finest-scale structure of the PSF, we evaluate the PSF and reconstruct an image using 0.05 arcsecond pixels. A three-component noise model is used to account for independent doses of readout noise, Poisson noise, and flat-field noise in the CCD data. A σ-clipping algorithm is implemented to find and reject corrupted data in pixels that were struck by cosmic rays during the exposure.

2 MEM Deconvolution of R136

30 Doradus is a spectacular region of star formation in the Large Magellenic Cloud, a dwarf companion of our Galaxy. The R136 star cluster in 30 Doradus contains recently-formed blue supergiant stars among which may be examples of the most massive stars that can form anywhere. In R136 we may be witnessing the birth of a globular cluster.

R136 was imaged with HST's Wide Field Camera for 40s through the F368M filter to support target acquisition for the first-light observations with the Goddard High-Resolution Spectrograph. While better data will be taken later in the mission, this first image of R136 has been a useful test case for deconvolution algorithms.

Figures 1 and 2 show two versions of the WFC image of R136 before and after MEM deconvolution. The brighter stars in the cluster core are individually visible in Figure 1, while Figure 2 shows the same data with a different grey-scale level which brings out hundreds of fainter stars while blacking out the cluster center. A bias level, determined in the usual way from unexposed parts of the CCD frame, has been subtracted from the raw data frame. Flat-field corrections have also been applied, but these are based on the F336W filter since flat field were not yet available for F368M. Note that the halos around individual stars and the fog of overlapping halos pervading the crowded star field at the cluster center are effectively removed by MEM deconvolution.

MEM's reconstructed image, when convolved with the PSF, is forced to fit the observed data to within $\chi^2/N = 1$, where N is the number of data pixels. Figure 3 compares the observed data with the computed data, obtained by convolving the reconstructed image with PSF. The computed data are slightly smoother than the observed data, since MEM finds the "smoothest" image that fits the observed data. This suppresses noise, but also entails some loss of resolution and biases the fluxes of point sources downward by about 1-σ. Note that several cosmic ray features in the observed data were flagged as bad data and hence are not modelled by the convolution.

3   MEM Deconvolution

MEM is a general tool for fitting a model to observational data which depend in a known way on an unknown positive additive image. In the deconvolution problem, the positive additive image we seek is the photon flux distribution on the plane of the sky. We must first develop a suitable model to describe the mapping from an image of the sky to observed data. This mapping may involve convolving the image with a point-spread function, binning it onto detector pixels, multiplying it by detector response, adding to it detector backgrounds, and any other steps deemed necessary. Here MEM affords useful flexibility to account for field distortions, position-dependent point-spread functions, non-linear responses, position-dependent and signal-dependent noise, and other peculiarities of the observational equipment.

Once a model of the imaging experiment has been defined, the MEM technique is to adjust the input image until it fits the data and is as simple as possible, where by simple we mean the absence of structure as measured by the image entropy, defined below. Given an input image, the model is used to compute predicted data values and a χ^2 statistic is then computed to measure how well the input image fits the data. We require require $\chi^2/N = 1$, but there are generally many different images that satisfy this data constraint, and so we need a regularizing condition to make the solution unique. The MEM solution is to maximize the image entropy subject to the constraint $\chi^2/N = 1$. The entropy, given by

$$S(f, m) = \sum_i f_i - m_i - f_i \ln(f_i/m_i),$$

measures of how close the image f is to the default image m.

The default image m is used to incorporate prior information. The entropy defined above is maximized when $f_i = m_i$, thus any pixels not constrained by data will default to m_i. The observations pull image values away from their default levels, while the entropy pulls the image back toward the default image. If no prior information is available, m can be set to a uniform value, for example the mean value of the image f. In that case MEM finds the "most uniform" image that fits the data. The reconstructions shown in this paper use the so-called "curvature" default, with m set equal to a slightly blurred version of the image f. In this case, S is a global measure of the image curvature, and MEM delivers the "smoothest" image that fits the data.

Iterative adjustment of the image to fit the data and maximize the entropy was accomplished using a 1985 version of the general MEM fitting package MEMSYS, kindly made available by Gull. This required coding two subroutines, OPUS and TROPUS, to perform the mapping from image to data space (OPUS) and the transpose mapping from data to image space (TROPUS). The algorithm is the Historic MEM documented by Skilling in this proceedings.

4   PSF Notes

The PSF's sharp core is an in-focus image formed by zones near the center of the mirror, while the superposition of out-of-focus images from other zones produces the broad halo. Additional fine-scale structure described as "hairy tendrils" radiating out from and encircling the core are diffraction effects from pupil obscurations. These include the secondary mirror and its four spider vanes, three circular "pads" on the front surface of the primary mirror, the the secondary mirror of the WFC Cassegrain repeater optics, and its 3-vaned spider support.

In this reconstruction experiment we used an empirical PSF derived using DAOPHOT from the images of three bright and fairly isolated stars located outside the sub-image. The PSF was evaluated on a 21×21 grid of 0.05 arcsecond pixels. We assume that a single PSF applies throughout the field of view. This is not a bad approximation for the 100 pixel sub-image we consider here.

In fact the shape of the PSF varies significantly with position because, as we move around in the field of view, pupil obscurations due to WFC Cassegrain repeater secondary mirror and spider translate with respect to the pupil of the primary mirror. Another complication, especially for wide-band filters, is the PSF's wavelength dependence. Details of the PSF depend on the unknown color of the sources being imaged. Such effects can in principle be treated in the MEM formulation, and will need to be tackled if the full information content of the data is to be recovered, but they are ignored in the present implementation.

Our model mapping the image $N(k)$ to the data $D(i)$ is given by

$$D(i) = \sum_j PIX(i,j) \sum_k PSF(j,k)N(k).$$

Here $N(k)$ is the distribution of photons detected during the exposure from image pixel k. The point-spread function $PSF(j,k)$ gives the fraction of the photons from pixel k that get moved to pixel j. $PIX(i,j)$ gives the fraction of image pixel j that goes into data pixel i. Each data pixel corresponds to a block of four image pixels, thus the sum over j simply adds up the image pixels in blocks of 4 to obtain the final data pixel i. Note that the image pixels extend outside the data frame by half the PSF width, so that the PSF convolution can be computed for data pixels on the edge of the data frame.

5   3-component Noise Model for CCD Data

In modelling the CCD data, we consider 3 independent sources of noise. First, readout noise with a standard deviation σ_0 is assumed to be present in every pixel. Poisson noise, due to the finite number of detected photons, is proportional to the square root of the detected signal. Finally, because the flat field corrections were not perfect, we assume a third independent noise which is linearly proportional to the signal. The variances of these three independent noise sources are added in quadrature. Thus the variance is given by

$$\sigma^2(D(i)) = \sigma_0^2 + D(i)/G + f^2 D(i)^2.$$

The three noise model parameters in this experiment are the rms readout noise $\sigma_0 = 2.4$ data numbers, the gain $G = 7.5$ photons per data number, and the fractional flat-field uncertainty $f = 0.03$.

To compute χ^2, we need to have variance estimates for each data point. Normal practice would assign error bars once and for all to the (noisy) observed data. However, one should more properly compute the error bars from the (noise-free) computed data obtained by convolving the image with the PSF. We use the above equation to re-compute the error bars from the computed data after each iterative adjustment of the image.

6   Cosmic Ray Rejection

A difficult problem in the analysis of CCD images is the treatment of data corrupted by cosmic ray hits on the detector. In WF/PC data these hits range in appearance from single high pixels to long streaks of high pixels, depending on the energy and angle of incidence of the particle track through the CCD. With MEM we can simply flag the bad pixels by assigning them very large error bars so that they are ignored in the calculation of χ^2.

A sigma-clipping algorithm was implemented in an attempt to automatically find and reject the cosmic ray hits. The search for bad pixels was performed between successive MEMSYS iterations. After predicted data values and error bars are computed, pixels were flagged as unreliable if the observed value exceeded the predicted value by more than $K\sigma$. Note in the figures that cosmic ray hits at $(X,Y) = (94,2)$ and $(80,59)$ were successfully flagged. These pixels have high values, but they lack the broad wings which would be present due to the PSF if they were real stars.

However, the cosmic ray rejection algorithm was not fully automatic. We had to be quite conservative by setting K to 10 during the early iterations in order not to incorrectly reject pixels near the cores of bright stars. After a reasonably low value of χ^2 was reached, we reduced K to 6. Further experimentation would be needed to see if a fully satisfactory automatic procedure can be identified.
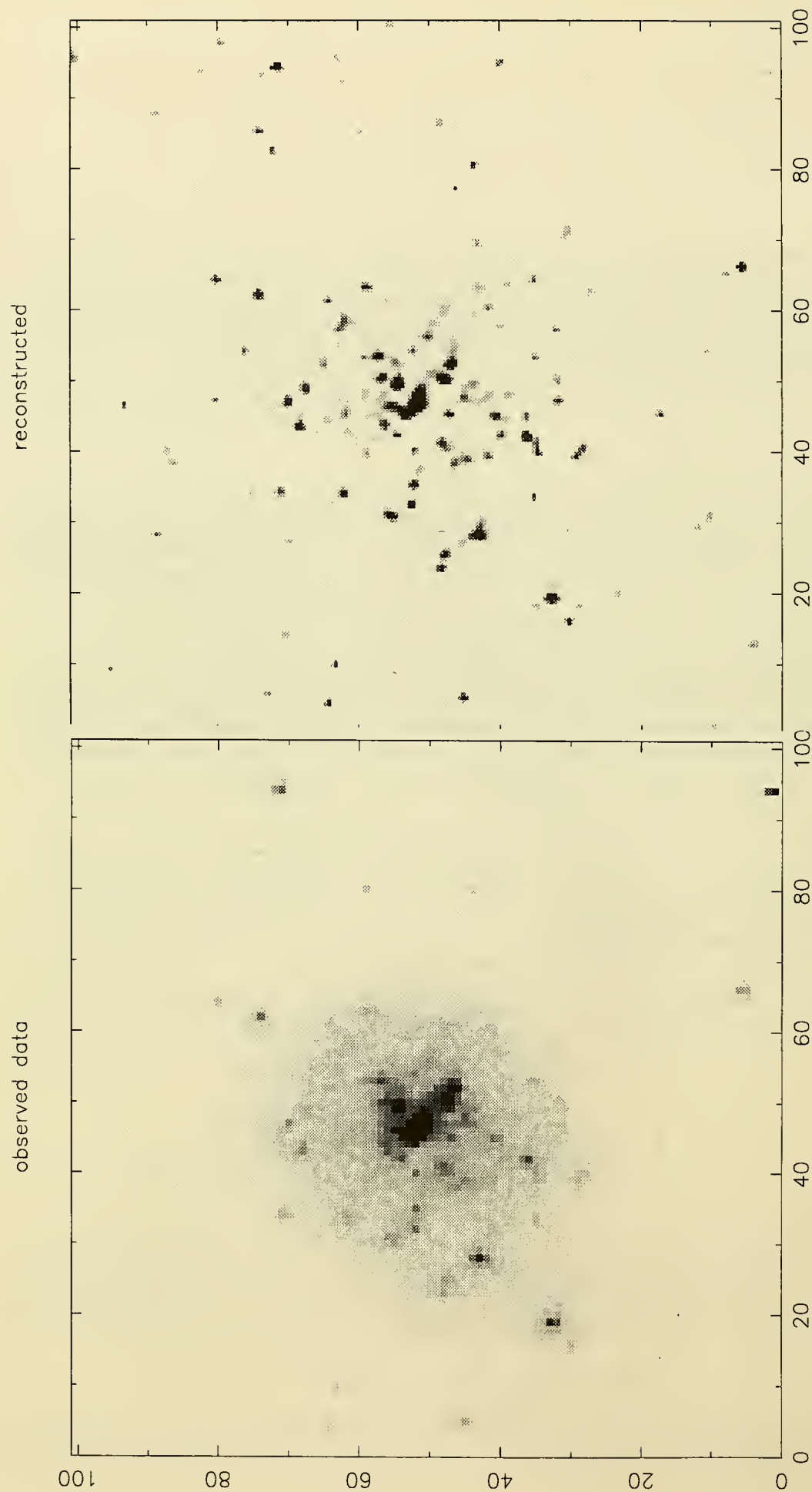
reconstructed

observed data

Figure 1 - Observed WFC image of R136 (left) and reconstructed image (right) obtained after MEM deconvolution. The pixels are 0.1×0.1 arcseconds in the observed image, and 0.05×0.05 arcseconds in the reconstructed image. Note that cosmic ray hits located near $(X,Y) = (94,2)$ and $(80,59)$ were rejected and therefore are absent in the reconstructed image.
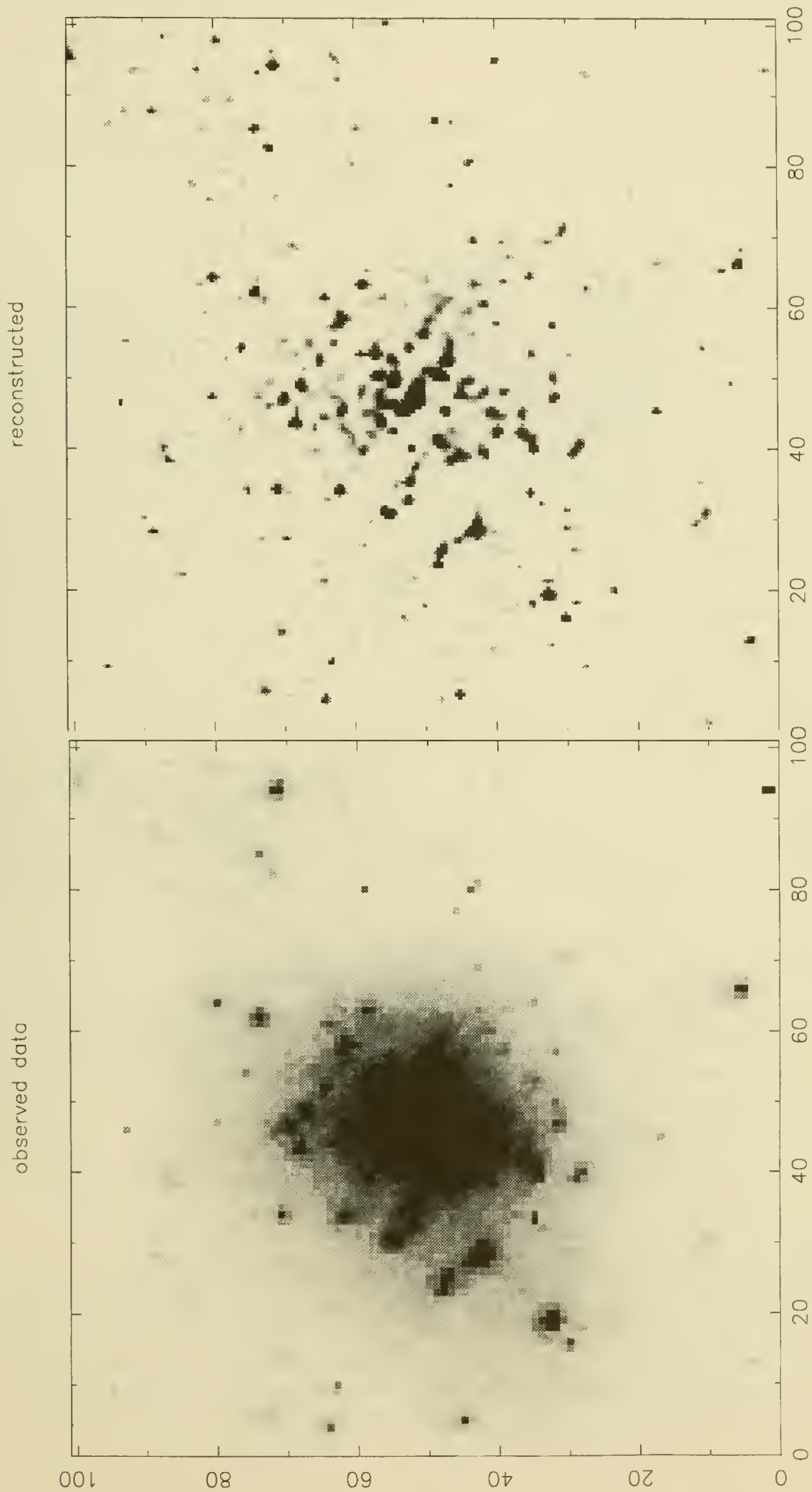
reconstructed

observed data

Figure 2 - Same as Figure 1 but with grey-scale level set to reveal hundreds of fainter stars.
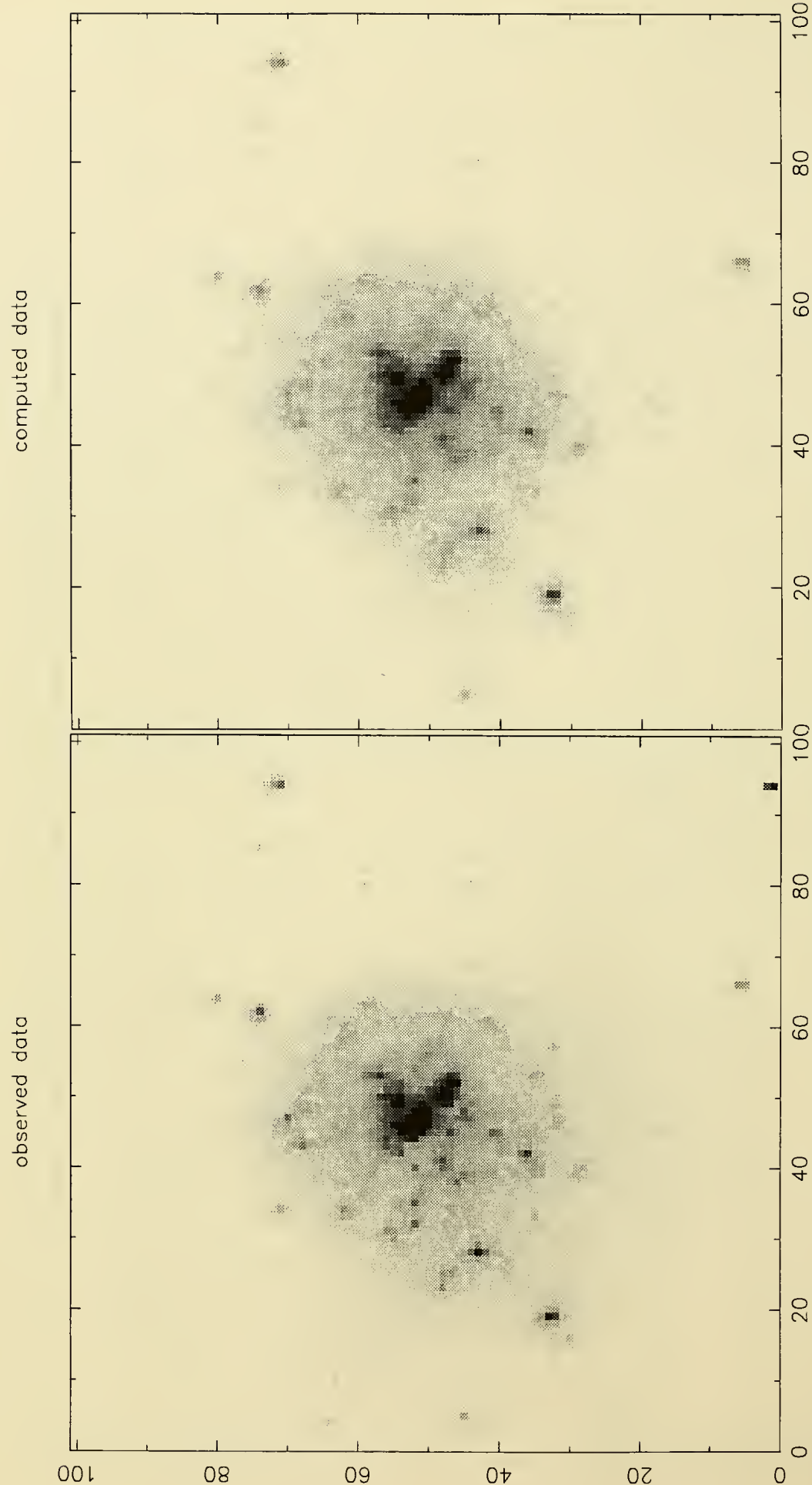
137

Figure 3 - The observed data are compared with computed data obtained by convolving the reconstructed image by the point-spread function and binning the resulting 0.05 arcsecond pixels in blocks of 4 to correspond with the 0.1 arcsecond data pixels. Note that cosmic ray hits located near $(X,Y) = (94,2)$ and $(80,59)$ were rejected and therefore are absent in the computed data.

Restoration of HST Images Using the Lucy Method with Oversampling

Richard L. White, STScI

Introduction

I have been doing a number of experiments in the restoration of Hubble Space Telescope (HST) images using the Lucy method (Lucy 1974). Since this method is well described elsewhere in this volume (see the paper by Snyder), in this short paper I describe only the method used to implement oversampling and show some results.

Lucy and Baade (1989; see also Lucy in this volume) discuss the advantages of restoring an image to a finer grid of pixels than the observed image. For example, in the results shown below a Wide Field Camera (WFC) image with 0.1" pixels has been deconvolved using 0.05" pixels, so that there are 4 pixels in the deconvolved image for each pixel in the observed image. This technique is sure to be of great importance in the analysis of WFC data, because the WFC grossly undersamples the HST point-spread function (PSF). The FWHM of the core of the HST PSF is considerably smaller than 0.1"; consequently the appearance of a star observed with the WFC can change dramatically as the telescope pointing is shifted by a fraction of a pixel. This is very troublesome to ordinary deconvolution methods that restore the image on the same grid as the observed data.

Any deconvolution technique that draws a clear distinction between model space and data space allows the possibility of using a finer grid for the model image than for the observed data. For example, Djorgovski and Weir (in this volume) use the maximum entropy method with oversampling, and Lucy and Baade (1989) use the Lucy method with oversampling. On the other hand, methods such as optimal filtering (Hanisch, this volume) and Jansson's method (Gilliland, this volume) do not permit a simple extension to oversampled grids; this must be considered a mark against such techniques.

Oversampling with the Lucy Method

Lucy and Baade (1989) show some results for simulated images using the Lucy method with oversampling. They do not, however, describe in any detail the form the iteration takes when oversampling is included, so it is worthwhile to describe it here. The usual form of the Lucy iteration is (Lucy 1974)

$$\phi(x) = \int \psi^r(\xi) P(x|\xi) d\xi \quad , \text{ and}$$

$$\psi^{r+1}(\xi) = \psi^r(\xi) \int \frac{\tilde{\phi}(x)}{\phi(x)} P(x|\xi) dx \quad ,$$

where $\tilde{\phi}(x)$ is the observed image, $\psi^r(\xi)$ is the model image at iteration r, and $P(x|\xi)$ is the PSF. Note that the PSF need not be shift-invariant, though if it is — in which case $P(x|\xi) = P(x - \xi)$ — the integrals become simple convolutions and can be done with fast Fourier transforms.

Suppose now that we have discrete images $\tilde{\phi}_i$, $i = 0, N - 1$ and ψ_k^r, $k = 0, BN - 1$, where B is the oversampling factor. If the original pixels are of size s, then the oversampled

pixels are of size s/B. As Lucy and Baade point out, the PSF must also be known on a fine grid with pixels of size s/B. They use the PSF tabulated on a fine grid but averaged over pixels of size s; I prefer to think of the PSF tabulated on the fine grid and averaged over the fine pixels as well, so that the finely sampled PSF is simply what one would observe using a detector that has B times better resolution than the real detector. Then the integrals become sums, and with a little algebra we find

$$d_l = \sum_{k=0}^{BN-1} \psi_k^r P_{lk} \quad , \quad l = 0, BN - 1,$$

$$\phi_i = \sum_{b=0}^{B-1} d_{b+Bi} \quad , \quad i = 0, N - 1,$$

$$R_{b+Bi} = \frac{\tilde{\phi}_i}{\phi_i} \quad , \quad b = 0, B - 1 \quad , \quad i = 0, N - 1 \text{ and}$$

$$\psi_k^{r+1} = \psi_k^r \sum_{l=0}^{BN-1} R_l P_{lk} \quad , \quad k = 0, BN - 1$$

Here for clarity I've introduced the variables d and R. The steps are these:

(1) Blur the model image ψ with the PSF to get d, the blurred image on the oversampled grid.

(2) Bin d in blocks of size B to get ϕ, the blurred image on the observed grid.

(3) Take the ratio of $\tilde{\phi}/\phi$ on the observed grid and use simple pixel replication to expand it by a factor B, giving R, the ratio on the oversampled grid.

(4) Convolve R with the transpose of the PSF and multiply by ψ^r to get ψ^{r+1}, the next iterate.

The first and fourth steps are exactly equivalent to the usual Lucy iteration, and they can be computed using FFTs if the PSF is shift-invariant. The second and third steps are also easily computed. The compute time for the entire calculation is dominated by the convolutions, which scale as $BN \log BN$ when FFTs can be used.

A slightly different approach allows the computation to be done in a computing time that scales $BN \log N$, which is slightly faster than the calculation outlined above. However, the above calculation is somewhat more flexible (*e.g.*, it permits a fully shift-variant PSF and makes some regularization methods easier to implement); the price in additional computing for this flexibility is typically small (10–20%).

The extension of this method to two-dimensional images is obvious and straightforward; the verbal description given above applies exactly to the higher dimensional case. Snyder (this volume) discusses various modifications of this iteration to accomodate dark current, flat field variations, etc.; all such modifications are easy to include with oversampling as well.
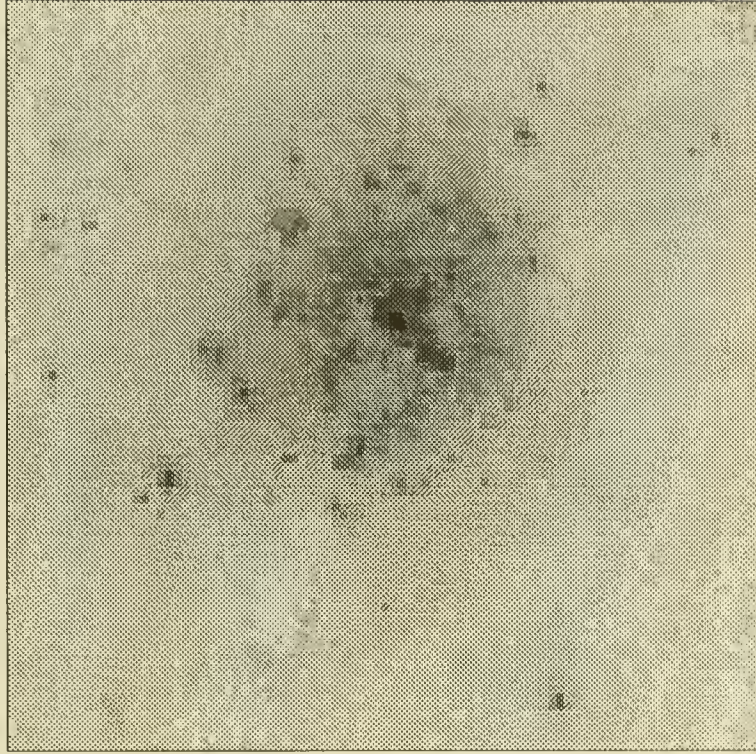
Figure 1: Raw WFC image of R136.

Application to HST Data

Figure 1 shows the WFC image of the central region of the R136 star cluster in the 30 Doradus region of the Large Magellanic Cloud. The portion of the image shown is 90 pixels (9") on a side; north is at the top and east is to the right. This image (and those in Figures 2 and 3) is displayed with a logarithmic grey scale which emphasizes structure near the sky noise level. Of course the images shown here do not show the full range of the data, but they are sufficient for the purposes of this paper.

Figure 2 shows the result of 120 Lucy iterations using a grid that is not oversampled. Snyder's modification for readout noise (described in this volume) has been used, which considerably reduces the amplification of noise in the sky. The star images here are, like those in the original data, rather different from one star to another due to the undersampling, and it is difficult to determine whether extended objects are a blend of two stars or whether they might be the result of truly diffuse emission.

Figure 3 displays the result of 120 Lucy iterations using a grid that is oversampled by a factor of $B = 2$ in each direction. Here the results are remarkably improved over Figure 2. The images of stars are much rounder and are more consistent from one star to another, and some of the ambiguous double stars are clearly separated. Note, for example, the double star about 2.5" due east of the brightest central star; it is very well-resolved in Figure 3, but is a somewhat amorphous blob in both Figures 1 and 2.

Another important feature of restoring WFC images using oversampled grids is that a significantly better fit to the data is obtained. If the signal-to-noise ratio of the data is
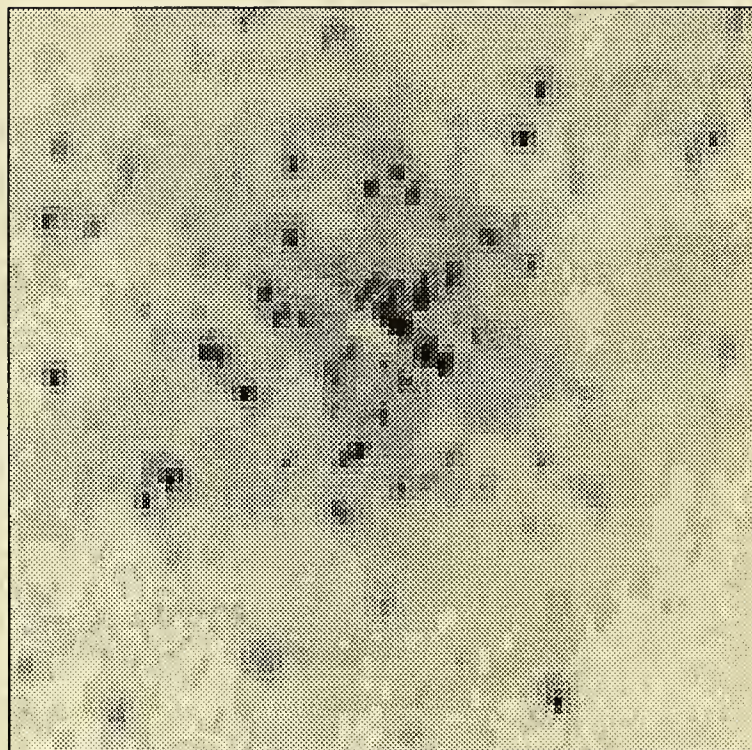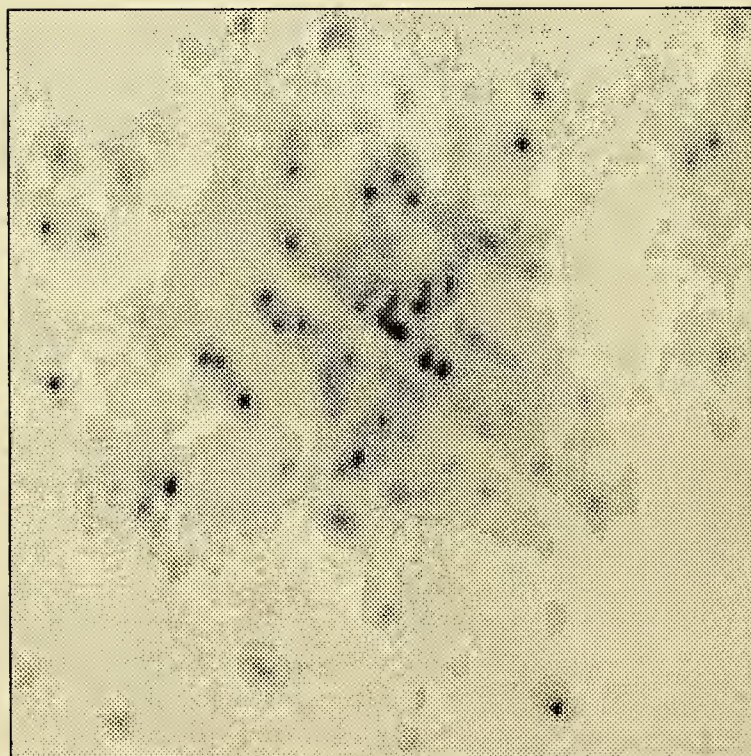
Figure 2: *R136 image restored with Lucy method.*

Figure 3: *R136 image restored with Lucy method using 2 × 2 oversampling.*

high, it is impossible to find an accurate fit of the undersampled model to the data because all stars are effectively forced to be in the centers of pixels. A star which is not centered on a pixel must necessarily be restored as a combination of a centered point source and some extended flux which compensates for the centering error; this limits the photometric fidelity of any restoration method. Restoring to a finer grid allows more of the flux in stellar halos to be put back into the center of the star image where it belongs.
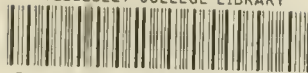
Conclusions

The restoration of images on a finer grid than the observed image has many advantages. For data taken with the HST Wide Field Camera, the most important advantage is that the finer grid allows one to compensate as far as is possible for the undersampling of the PSF by the camera. Oversampling can be implemented using any deconvolution algorithm which distinguishes the object and image spaces; the Lucy method is one such algorithm which is well-suited to HST data. The addition of oversampling to the Lucy algorithm does not significantly complicate the implementation of the method. The computing time required for an oversampled image with $B \times B$ pixels per camera pixel is about B^2 times as large as that for restoration matching the camera resolution; however, for demanding applications this extra computing time gives improved results which are well worth the effort.

More work must be done to establish the usefulness of images generated using this (or any other) deconvolution method for quantitative analysis, but initial experiments with simulated data are encouraging. Other work in progress includes the use of two or more images shifted by a fraction of a pixel to construct a single model image; this application also obviously requires the restored image to be on a finer grid than any of the observed images.

References

Lucy, L. B. 1974, *A. J.*, **79**, 745.
Lucy, L. B., and Baade, D. 1989, in *Proc. 1st ESO/ST-ECF Data Analysis Workshop, Garching, April 1989*, eds. P. Grosbøl, R. H. Warmels, F. Murtagh, p. 219.

DATE DUE